



دانشکده‌ی مهندسی کامپیوتر و فن‌آوری اطلاعات
گروه هوش مصنوعی

سیستم پیشنهادگر هوشمند جهت تعیین افراد تاثیرگذار در شبکه‌ی اجتماعی

دانشجو: بهزاد رضایی

استاد راهنما:

دکتر مرتضی زاهدی

استاد مشاور:

دکتر هدی مشایخی

پایان‌نامه‌ی ارشد جهت اخذ درجه‌ی کارشناسی ارشد

بهمن ماه ۱۳۹۴

دانشگاه صنعتی شاهرود
 دانشکده: مهندسی کامپیوتر و فن آوری اطلاعات
 گروه: هوش مصنوعی
 پایان نامه کارشناسی ارشد آقای بهزاد رضایی
 تحت عنوان: سیستم پیشنهادگر هوشمند جهت تعیین افراد تاثیرگذار در شبکه‌ی اجتماعی

در تاریخ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد
 مورد ارزیابی و با درجه مورد پذیرش قرار گرفت.

امضا	اساتید مشاور	امضا	اساتید راهنما
	نام و نام خانوادگی:		نام و نام خانوادگی:
	نام و نام خانوادگی:		نام و نام خانوادگی:

امضا	نماینده تحصیلات تکمیلی	امضا	اساتید داور
	نام و نام خانوادگی:		نام و نام خانوادگی:
			نام و نام خانوادگی:
			نام و نام خانوادگی:
			نام و نام خانوادگی:

تقدیم بہ:

پدر و مادر عزیز و فداکارم

ہمراہ ہمیشگی م، ہمسر مہربانم

شکر و قدردانی

«من لم یشکر المخلوق، لم یشکر الخالق»

از اساتید راهنا و مشاور خود به ترتیب، آقای دکتر مرتضی زاهدی و خانم دکتر مهدی مشایخی که همواره بهترین راهنمایی ها و مشاوره ها را در تک تک مراحل انجام پژوهش حاضر داشته اند، قدردانی می کنم. همچنین از دیگر اساتید محترم و با فضیلت خویش، جناب آقایان دکتر حمید حسن پور، دکتر علی اکبر پویان و دکتر صادق رحیمی شهربان مقدس که همواره در طول تحصیل، اینجانب را راهنمایی و حمایت کرده اند تا بتوانم در مسیر صحیح علم و دانش گام بردارم، شکر می نمایم.

سپاس گذاری می کنم از کلیه اعضای هیئت علمی دانشکده ی مهندسی کامپیوتر و فن آوری اطلاعات و کارشناسان و نشی محترم که هر یک به نوبه ی خود اینجانب را در مسیر آموزش و پیشرفت هر چه بهتر و پیش تر کمک کردند.

از مدیریت محترم شرکت تدبیر پرداز با بابت همکاری شایان تقدیرشان در مدیریت کارهای محوله به اینجانب به منظور تسهیل در فرآیند انجام پایان نامه تقدیر و شکر می کنم.

در پایان بر خود لازم می دانم از دوست عزیزم، جناب آقای مهندس علی رضامانشتی جهت فراهم نمودن امکانات لازم به منظور تسهیل در فرآیند جمع آوری داده ها شکر و قدردانی نمایم.

تعهد نامه

اینجانب بهزاد رضایی دانشجوی دوره کارشناسی ارشد رشته مهندسی کامپیوتر، گرایش هوش مصنوعی دانشکده مهندسی کامپیوتر و فن آوری اطلاعات دانشگاه صنعتی شاهرود نویسنده پایان نامه سیستم پیشنهادگر هوشمند جهت تعیین افراد تاثیرگذار در شبکه‌ی اجتماعی تحت راهنمایی دکتر مرتضی زاهدی متعهد می شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

* متن این صفحه نیز باید در ابتدای نسخه های تکثیر شده پایان نامه وجود داشته باشد .

چکیده

با توجه به روند رو به رشد شبکه‌های اجتماعی و استفاده از آن‌ها به عنوان یکی از مهم‌ترین ابزارهای ارتباطی در دنیای امروز، تحلیل آن‌ها از اهمیت ویژه‌ای برخوردار است. از جمله کاربردی‌ترین تحلیل‌ها، تشخیص تأثیرگذارترین افراد در این شبکه‌ها است. افراد تأثیرگذار در یک شبکه‌ی اجتماعی به کاربرانی اطلاق می‌شود که می‌توانند با رفتار خود در این شبکه‌ها، نظرها و عقاید دیگر کاربران را تغییر دهند. این نوع تحلیل در زمینه‌های مختلفی از جمله تجارت، بازاریابی و سیاست پر کاربرد است. از آنجایی که ویژگی‌های شبکه‌های اجتماعی مختلف با یکدیگر متفاوت است، تعیین افراد تأثیرگذار در هر یک از آن‌ها نیز متفاوت است. هدف پژوهش حاضر، ارائه‌ی یک روش جدید در جهت تشخیص افراد تأثیرگذار در شبکه‌ی اجتماعی توئیت^۱ است.

در این تحقیق، با جمع‌آوری داده‌های مورد نیاز از شبکه‌ی اجتماعی توئیت در رابطه با حملات تروریستی سال ۲۰۱۵ شهر پاریس، مدلی برای تشخیص افراد تأثیرگذار ارائه شده است. در مدل پیشنهادی، با استفاده از ویژگی‌های موجود در هر توئیت و نیز ویژگی‌های کاربر منتشرکننده‌ی آن که از طریق رابط کاربردی برنامه‌نویسی^۱ توئیت قابل دسترسی است، افراد تأثیرگذار تشخیص داده می‌شوند. در این روش، ویژگی‌ها به چهار گروه تقسیم‌بندی می‌شوند: (۱) ویژگی‌هایی از کاربران که از تعامل با دیگر افراد مشخص می‌شوند؛ (۲) ویژگی‌هایی از کاربران که از طریق پروفایلشان تعیین می‌شوند؛ (۳) ویژگی‌هایی از توئیت‌ها که دیگر کاربران آن‌ها را مشخص می‌کنند؛ (۴) ویژگی‌هایی از توئیت‌ها که در خود توئیت قرار دارد. پس از انجام فرآیند پیش‌پردازش بر روی داده‌های جمع‌آوری شده، وزن اختصاصی ویژگی‌های مرتبط با کاربران و توئیت‌ها، به طور جداگانه و با استفاده از روش AHP محاسبه می‌شود. در قسمت دیگری از فرآیند تشخیص افراد تأثیرگذار، توئیت‌های موجود خوشه‌بندی شده و رابطه‌ی معنایی بین توئیت‌های هر خوشه مورد بررسی قرار می‌گیرند. در نهایت، پس از دخیل کردن عامل زمان به عنوان یکی از مهم‌ترین عوامل در تغییر میزان تأثیرگذاری، مدل مورد نظر ساخته و در پایگاه داده‌ی جمع‌آوری شده اعمال می‌شود.

نتایج به دست آمده حاکی از اهمیت امتیاز توئیت‌ها نسبت به امتیاز کاربری است. هم‌بستگی بین رتبه‌ی کاربران و میانگین امتیاز هر توئیت عدد ۰٫۶۱۹ را نشان می‌دهد. این عدد بیان‌گر تأثیر مستقیم کیفیت توئیت‌ها بر رتبه‌ی کاربران است. همچنین نتایج نشان می‌دهد کیفیت پیام توئیت‌ها از تعداد

^۱ Application Programming Interface (API)

توئیت‌ها پراهمیت‌تر است؛ به عبارت دیگر افزایش تعداد توئیت‌ها، نمی‌تواند بالا بردن میزان تاثیرگذاری کاربر را تضمین کند.

روند تاثیرگذاری کاربران در بازه‌ی ۲۱ روزه‌ی جمع‌آوری داده‌ها نشان می‌دهد ۷۵ درصد از افرادی که در هر روز جزو تاثیرگذارترین‌ها دسته‌بندی می‌شوند، نمی‌توانند روند تاثیرگذاری خود را در روزهای دیگر حفظ کنند. همچنین ۵۹ درصد از افرادی که در هر روز رتبه‌های اول تا سوم را کسب می‌کنند، به همان اندازه که سریع به این جایگاه‌ها می‌رسند، با همان سرعت نیز آن را از دست می‌دهند. هشتگ‌های مورد استفاده در پنج توئیت تاثیرگذار در هر روز نشان‌دهنده‌ی جایگاه ۲۰ درصدی هشتگ #prayforparis می‌باشد. در رده‌ی دوم، هشتگ #isis با ۳ درصد قرار می‌گیرد.

هم‌بستگی بین متغیرهای در نظر گرفته شده برای کاربران و نتایج به دست آمده برای امتیاز آن‌ها، نشان دهنده‌ی رابطه‌ی قوی ۰,۹۹ بین تعداد دنبال‌کنندگان و امتیاز یک کاربر است. این در حالی است که بیش‌ترین هم‌بستگی بین مغیرهای هر توئیت و امتیاز توئیت‌ها، ۰,۳۷ است که این عدد مربوط به ویژگی تعداد دفعاتی است که یک توئیت، مورد پسند دیگران واقع شده است.

کلمات کلیدی: کاربران تاثیرگذار توئیت‌ر، تحلیل شبکه‌های اجتماعی، سیستم پیشنهادگر

مقالات مستخرج از پایان نامه

1. Behzad Rezaie, Morteza Zahedi, Hoda Mashayekhi, “Content and Network Analysis of Iranian Sports Weblogs”, Journal of Soft Computing and Information Technology (Submitted).

فهرست مطالب

۱- مقدمه.....	۱
۱-۱- پیش‌گفتار.....	۲
۲-۱- شبکه‌های اجتماعی.....	۳
۱-۲-۱- شبکه‌های اجتماعی ارتباط عمومی.....	۳
۲-۲-۱- شبکه‌های اجتماعی اشتراک‌گذاری چندرسانه‌ای.....	۳
۳-۲-۱- شبکه‌های اجتماعی حرفه‌گرا.....	۴
۴-۲-۱- شبکه‌های اجتماعی انتشار مطلب.....	۴
۵-۲-۱- شبکه‌های اجتماعی انجمنی.....	۵
۶-۲-۱- شبکه‌های اجتماعی پیام‌رسان.....	۵
۳-۱- تعریف مسئله و اهمیت آن.....	۵
۴-۱- چالش‌ها.....	۶
۲- مروری بر کارهای پیشین.....	۹
۱-۲- سنجه‌های مرکزیت در گراف.....	۱۰
۱-۱-۲- سنجه‌ی لینک ورودی.....	۱۰
۲-۱-۲- سنجه‌ی بینابینی.....	۱۰
۳-۱-۲- سنجه‌ی نزدیکی.....	۱۱
۲-۲- مدل‌های خاص.....	۱۲
۱-۲-۲- وبلاگ‌ها.....	۱۲
۲-۲-۲- توئیتر.....	۲۳
۳-۲- مدل‌های عمومی.....	۳۳
۴-۲- بررسی ابزارهای رتبه‌بندی و تحلیلی.....	۳۶
۱-۴-۲- Klout.....	۳۷
۲-۴-۲- Twitalyzer.....	۳۸
۳-۴-۲- Twittercounter.....	۳۸
۴-۴-۲- Hootsuite.....	۴۰
۳- روش پیشنهادی تعیین افراد تاثیرگذار در شبکه اجتماعی.....	۴۵
۱-۳- معرفی توئیتر.....	۴۶
۱-۱-۳- توئیٹ.....	۴۶
۲-۱-۳- دنبال‌کننده.....	۴۶
۳-۱-۳- هشتگ.....	۴۷

۴۸	ذکر نام کاربران	۴-۱-۳
۴۸	بازنشر توئیت و مورد پسند واقع شدن	۵-۱-۳
۴۸	رابط کاربردی برنامه‌نویسی توئیتر	۶-۱-۳
۴۹	ابزارهای مورد استفاده در پژوهش	۲-۳
۴۹	رابط کاربردی برنامه‌نویسی Google Maps Geocoding	۱-۲-۳
۵۰	نرم‌افزار Carrot	۲-۲-۳
۵۶	کتابخانه‌ی Document Similarity	۳-۲-۳
۵۷	نرم‌افزار Gephi	۴-۲-۳
۵۹	کتابخانه‌ی Twitter4j	۵-۲-۳
۶۰	پایگاه داده	۳-۳
۶۰	موضوع داده‌های جمع‌آوری شده	۱-۳-۳
۶۰	نحوه‌ی جمع‌آوری داده‌ها	۲-۳-۳
۶۲	پارامترهای جمع‌آوری شده در داده‌ها	۳-۳-۳
۶۴	اطلاعات آماری پایگاه داده‌ی جمع‌آوری شده	۴-۳-۳
۶۶	بستر پیاده‌سازی	۴-۳
۶۷	پیش‌پردازش داده‌ها	۵-۳
۶۷	حذف هرزنامه‌نویسان	۱-۵-۳
۶۸	حذف هشتگ‌ها، نام کاربران و لینک‌های آدرس	۲-۵-۳
۶۸	حذف توئیتهای با شباهت کم معنایی	۳-۵-۳
۶۹	پارامترها	۶-۳
۷۰	تنظیم پارامترها	۷-۳
۷۱	تعیین وزن پارامترها	۱-۷-۳
۷۶	تعیین میزان شباهت توئیتهای	۲-۷-۳
۸۱	تعیین فاصله‌ی بین محل انتشار توئیث و محل حادثه	۳-۷-۳
۸۴	معرفی مدل	۸-۳
۸۴	مدل تاثیرگذاری کاربران	۱-۸-۳
۸۵	مدل تاثیرگذاری توئیتهای	۲-۸-۳
۸۶	مدل جامع	۳-۸-۳
۸۷	نتیجه‌گیری	۹-۳

۴- نتایج و تحلیل داده‌ها..... ۸۹

۹۰	امتیازهای مدل‌های تاثیرگذاری	۱-۴
۹۰	مدل تاثیرگذاری کاربران	۱-۱-۴
۹۱	مدل تاثیرگذاری توئیتهای	۲-۱-۴
۹۴	مدل جامع	۳-۱-۴

۹۶.....	روند تاثیرگذاری	۲-۴ -
۹۸.....	هشتگ‌های توئیت‌های تاثیرگذار	۳-۴ -
۹۹.....	همبستگی بین متغیرهای موجود	۴-۴ -
۱۰۳.....	مقایسه‌ی روش پیشنهادی با دیگر مدل‌ها	۵-۴ -
۱۰۳.....	مقایسه با مدل‌های موجود در وبلاگ‌ها	۴-۵-۱ -
۱۰۴.....	مقایسه با مدل‌های موجود در توئیتر	۴-۵-۲ -
۱۰۷.....	نتیجه‌گیری و کارهای آینده	۵- -
۱۰۸.....	نتیجه‌گیری.....	۵-۱ -
۱۰۸.....	پیشنهادها برای کارهای آینده	۵-۲ -

فهرست شکل‌ها

- شکل ۱-۲- مثالی بصری از سنج‌های بینابینی ۱۱
- شکل ۲-۲- مثالی بصری از سنج‌های نزدیکی ۱۲
- شکل ۳-۲- نمودار افراد تاثیرگذار در یک بازه‌ی ۳ ساله با توجه به مدل *Influence Flow* [۷] ۱۵
- شکل ۴-۲- نمودار رسم شده بر اساس معیارهای *Topic Score* و *Geo Relevance Score* در [۳] ۲۶
- شکل ۵-۲- رتبه‌بندی افراد تاثیرگذار بر اساس تعداد *retweet*ها در سه موضوع مختلف [۱۳] ۲۹
- شکل ۶-۲- مثالی از چند توثیت که با نرم‌افزار *LIWS* در زیرگروه‌های "ناراحتی" و "عصبانیت" دسته‌بندی شده‌اند [۱۵] ۳۱
- شکل ۷-۲- (الف) مثالی از تغییر احساسات در ۵ متن در طول زمان؛ (ب) ایجاد مرز جداکننده‌ی اول؛ (ج) مرز جداکننده‌ی دوم [۱۵] ۳۱
- شکل ۸-۲- مثالی از یک شبکه برای تشریح مدل مرکزیت محلی [۲۱] ۳۵
- شکل ۹-۲- عدد تاثیرگذاری آقای دکتر ظریف، وزیر امور خارجه‌ی کشورمان، در دو سایت *Klout* (عدد ۶۳) و *Twitalyzer* (عدد ۲۲) ۳۷
- شکل ۱۰-۲- مثالی از روند رو به رشد تاثیرگذاری کاربری با نام کاربری *behzad.rezaie11* در شبکه‌ی اینستاگرام. نمودار توسط ابزار *Klout* رسم شده است. ۳۸
- شکل ۱۱-۲- نمونه‌ای از آمار کاربر *@Behzad28* ارائه شده در *Twittercounter* ۴۰
- شکل ۱۲-۲- نمونه‌ای از مدیریت چند حساب کاربری در شبکه‌های گوناگون توسط ابزار *Hootsuite* ۴۲
- شکل ۱۳-۲- نمونه‌ای از آمار ارائه شده توسط *Hootsuite* برای کاربر *@Behzad28* ۴۳
- شکل ۱۴-۲- ویژگی‌های مهم استخراج شده از ابزار *PeenIndex*، *Klout*، *TwitterGrader* و (الف) درخت تصمیم *C4.5* رسم شده برای ابزار *Klout*؛ (ب) درخت رگرسیون *REPTREE* رسم شده برای ابزار *PeerIndex*؛ (ج) درخت رگرسیون *REPTREE* رسم شده برای ابزار *TwitterGrader* [۲۶] ۴۴
- شکل ۱-۳- نمونه‌ای از ساختار ورودی فایل *XML* به نرم‌افزار *Carrot* ۵۱
- شکل ۲-۳- نمونه‌ای از ساختار خروجی فایل *XML* به نرم‌افزار *Carrot* ۵۳
- شکل ۳-۳- نمونه‌ای از عبارات باقاعده‌ی تعریف شده در *Carrot* ۵۶
- شکل ۴-۳- نمونه‌ای از شبکه‌ی بصری‌سازی شده در *Gephi* [۳۱] ۵۸
- شکل ۵-۳- نمونه‌ای از اطلاعات توثیت استخراج شده از کاربری با نام کاربری *CarnegieEndow* ۶۳
- شکل ۶-۳- نمونه‌ای از اطلاعات استخراج شده از کاربری با نام کاربری *joeeyrichardson* ۶۵
- شکل ۷-۳- نمایش بصری گراف ساخته شده از ارتباط معنایی بین توثیت‌ها ۸۲
- شکل ۸-۳- نمایش بصری گام‌های برداشته شده از جمع‌آوری داده‌ها تا ساخت مدل ۸۷
- شکل ۱-۴- هشتگ‌های مورد استفاده در پنج توثیت تاثیرگذار در هر روز به همراه تعداد آن‌ها ۹۹

فهرست جدول‌ها

جدول ۱-۲- آمار پست‌ها و لینک‌های ورودی پایگاه داده‌ی مورد استفاده در [۸]	۱۷
جدول ۲-۲- نتایج بدست آمده از روش‌های <i>MEIBI Influence Flow</i> و <i>MEIBIX</i> ، راست: <i>Influence Flow</i> ؛ وسط:	
<i>MEIBI</i> ؛ چپ: <i>MEIBIX</i> [۸]	۱۸
جدول ۳-۲- میزان همبستگی بین مدل‌های <i>MEIBI Influence Flow</i> و <i>MEIBIX</i> [۸]	۱۸
جدول ۴-۲- اطلاعات پایگاه‌داده‌ی استفاده شده در [۹]	۲۰
جدول ۵-۲- مقایسه‌ی مدل‌های <i>MEIBI</i> ، <i>BI</i> ، <i>BP</i> و <i>MEIBIX</i> [۹]	۲۰
جدول ۶-۲- نتایج به دست آمده از تأثیرگذارترین افراد پس از اعمال مدل آلفا [۲]	۲۴
جدول ۷-۲- مقایسه‌ی رتبه‌بندی گره‌ها با استفاده از سنج‌های محلی (<i>L</i>)، درجه (<i>D</i>)، نزدیکی (<i>C</i>) و بینایی (<i>B</i>) در شبکه‌های (از چپ به راست) وبلاگ‌ها، روترها، دانش‌مندان و ایمیل‌ها [۲۱]	۳۵
جدول ۱-۳- مقایسه‌ی ویژگی‌های الگوریتم‌های <i>Lingo</i> ، <i>K-means</i> و <i>STC</i> در نرم‌افزار <i>Carrot</i>	۵۵
جدول ۲-۳- اطلاعات آماری مربوط به پایگاه‌داده‌ی توئیت‌ها	۶۶
جدول ۳-۳- اطلاعات آماری مربوط به پایگاه‌داده‌ی کاربران	۶۶
جدول ۴-۳- میزان پراکندگی روابط استخراجی از توئیت‌ها با مشابهت بیش از ۶۰ درصد	۶۸
جدول ۵-۳- نسبت اهمیت ویژگی‌های هر کاربر در الگوریتم <i>AHP</i>	۷۱
جدول ۶-۳- نسبت اهمیت ویژگی‌های هر توئیت در الگوریتم <i>AHP</i>	۷۴
جدول ۷-۳- اوزان محاسبه شده برای پارامترهای هر کاربر توسط الگوریتم <i>AHP</i>	۷۵
جدول ۸-۳- اوزان محاسبه شده برای پارامترهای هر توئیت توسط الگوریتم <i>AHP</i>	۷۶
جدول ۹-۳- تعداد توئیت‌ها، عبارات کلیدی و تعداد روابط معنایی بین توئیت‌ها در هر خوشه با استفاده از الگوریتم <i>STC</i> در نرم-افزار <i>Carrot</i>	۷۷
جدول ۱۰-۳- آمار به دست آمده از گراف استخراج شده از رابطه معنایی بین توئیت‌ها	۸۲
جدول ۱۱-۳- بیش‌ترین و کم‌ترین فاصله‌ی به دست آمده از توئیت ارسالی تا شهر پاریس	۸۴
جدول ۱-۴- رتبه‌ی کاربران حاضر در پایگاه‌داده به همراه ویژگی‌های آن‌ها پس از اعمال مدل تأثیرگذاری کاربران	۹۰
جدول ۲-۴- رتبه‌ی توئیت‌های حاضر در پایگاه‌داده به همراه ویژگی‌های آن‌ها پس از اعمال مدل تأثیرگذاری توئیت‌ها	۹۳
جدول ۳-۴- رتبه‌ی کاربران بر اساس مدل جامع	۹۴
جدول ۴-۴- تعداد توئیت‌های کاربران در پایگاه‌داده و میانگین امتیاز هر توئیت	۹۵
جدول ۵-۴- رتبه‌ی کاربرانی که توانسته‌اند حداقل دو روز در بین ده فرد تأثیرگذار در هر روز باشند به همراه روزهای تأثیرگذاری آن‌ها	۹۶
جدول ۶-۴- درجه‌ی همبستگی ویژگی‌های در نظر گرفته شده و نتایج به دست آمده برای کاربران	۱۰۰
جدول ۷-۴- درجه‌ی همبستگی ویژگی‌های در نظر گرفته شده و نتایج به دست آمده برای توئیت‌ها	۱۰۲
جدول ۸-۴- مقایسه‌ی نتایج به دست آمده با اعمال مدل ارائه شده در [۱۶] و مدل ارائه شده در تحقیق حاضر	۱۰۶

١- مقدمه

۱-۱- پیش‌گفتار

انسان، موجودی اجتماعی است و نمی‌تواند به تنهایی و در انزوا زندگی کند. این بُعد اجتماعی به ناچار او را مجاب می‌کند که برای رسیدن به اهدافش با دیگران در تعامل باشد. تعاملات انسان‌ها در دنیای حقیقی باعث شده است بتوانند از یکدیگر تأثیر بپذیرند. تأثیری که افراد از هم نوعان خود می‌گیرند، می‌تواند روش، هدف و نحوه‌ی زندگی آن‌ها را تغییر دهد؛ گاه بهتر و گاه بدتر کند. تا قبل از پیدایش رایانه و اینترنت، تأثیرپذیری انسان‌ها از یکدیگر تنها با ارتباط حقیقی بین آن‌ها صورت می‌گرفت، اما امروزه تغییراتی که در زندگی افراد رخ می‌دهد، می‌تواند ناشی از تأثیری باشد که کاربران اینترنتی از هم‌نوعان مجازی خود می‌پذیرند. سهم شبکه‌های اجتماعی در این میان هر روز بیش‌تر و بیش‌تر از دیگر رسانه‌ها و حتی تعاملات حقیقی است.

“شبکه‌ی اجتماعی”، مفهومی است که پیش از طرح در عرصه فضای مجازی، در فضای واقعی جوامع نیز دارای مفهوم است. هدف از تأسیس این شبکه‌ها، ایجاد ارتباطی مجازی بین انسان‌ها است؛ ارتباطی که در دنیای حقیقی از آن به عناوین مختلف مانند خویشاوندی، دوستی و ... یاد می‌شود.

شبکه‌های اجتماعی مجازی تنها ابزارهای تکنولوژیکی جدیدی نیستند که امکانات جالب‌توجهی را در اختیار کاربران اینترنتی قرار داده باشند. شبکه‌های اجتماعی را فراتر از یک وب‌سایت معمولی می‌توان به عنوان رسانه‌های جدیدی در نظر گرفت که در ساختارهای اجتماعی، فرهنگی، اقتصادی و سیاسی تغییراتی ایجاد کرده‌اند. شبکه‌های اجتماعی از سویی به عنوان یکی از گونه‌های رسانه‌های اجتماعی، امکانات تعاملی قابل‌توجهی برای کاربران اینترنتی فراهم کرده‌اند و در افزایش مشارکت شهروندان در برخی فرآیندها موثر بوده‌اند. از سویی این شبکه‌ها با آسیب‌های گسترده‌ای در حوزه‌هایی از قبیل حریم خصوصی، اعتیاد مجازی، سوءاستفاده از کودکان، دزدی اطلاعات و هویت و مواردی این‌چنینی مواجه بوده‌اند. این شبکه‌ها جنبه‌های مختلفی از جمله ارتباطات سیاسی را متحول کرده‌اند. رویدادهای سیاسی سال‌های اخیر از انتخاب ریاست جمهوری آمریکا [۱] گرفته تا حملات تروریستی به شهرهای مختلف دنیا و انقلاب‌های اخیر کشورهای عربی [۲،۳]، همگی متأثر از فعالیت‌های کاربران شبکه‌های اجتماعی بودند.

۱-۲- شبکه‌های اجتماعی

شبکه‌های اجتماعی، سرویس‌های تحت وبی هستند که به کاربران اجازه می‌دهند (۱) پروفایل‌های شخصی بسازند؛ (۲) افرادی را انتخاب نموده و با آن‌ها در ارتباط باشند و (۳) مطالب، عکس‌ها و فیلم‌های خود را با افراد انتخاب‌شده، به اشتراک بگذارند [۴]. از جمله شبکه‌های اجتماعی معروف می‌توان به فیس-بوک^۲، توئیتر^۳ و تلگرام^۴ اشاره کرد. گسترش روزافزون این شبکه‌ها، باعث شده افراد بسیار زیادی در سرتاسر دنیا با یکدیگر در ارتباط باشند و در این شبکه‌ها سیر کنند، به طوری که در سه ماهه‌ی سوم سال ۲۰۱۵، ماهانه ۳۰۷ میلیون نفر در توئیتر فعال بوده‌اند^۵.

شبکه‌های اجتماعی را می‌توان از نظر کاربرد به گروه‌های زیر دسته‌بندی کرد:

۱-۲-۱- شبکه‌های اجتماعی ارتباطی عمومی^۶

در این دسته، شبکه‌هایی قرار می‌گیرند که هدف اصلی آن‌ها ایجاد ارتباط بین دوستان، آشنایان و ... است. این شبکه‌ها، تمام ارتباطات ما را در یک مکان، مثل Wall یا Timeline نگه می‌دارند. از جمله شبکه‌های معروف در این زمینه می‌توان به فیس-بوک، توئیتر و گوگل پلاس^۷ اشاره کرد. پس از محبوبیت این شبکه‌ها در بین کاربران ایرانی، شبکه‌های مختلفی نیز در کشور ایجاد شدند که از محبوب‌ترین آن‌ها می‌توان به کلوب^۸ اشاره کرد.

۱-۲-۲- شبکه‌های اجتماعی اشتراک‌گذاری چندرسانه‌ای^۹

در این نوع شبکه‌ها، کاربران می‌توانند به سادگی با ایجاد یک حساب کاربری، فایل‌های چندرسانه‌ای خود را با دیگران به اشتراک بگذارند. اینستاگرام^{۱۰}، فلیکر^{۱۱} و پیکاسا^{۱۲}

^۲ <https://facebook.com>

^۳ <https://twitter.com>

^۴ <https://telegram.org>

^۵ <http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>

^۶ General Relationship Social Networks

^۷ <https://plus.google.com>

^۸ <http://cloob.com>

^۹ Multimedia Sharing Social Networks

^{۱۰} <https://instagram.com>

^{۱۱} <https://flickr.com>

^{۱۲} <https://picasa.google.com>

به عنوان چندی از محبوب‌ترین شبکه‌ها در به اشتراک‌گذاری عکس در این گروه قرار می‌گیرند. در به اشتراک‌گذاری ویدئو، شبکه‌ی یوتیوب^{۱۳} گوی را از دیگر رقبای خود ربوده است. در ایران نیز، شبکه‌ی اجتماعی آپارات^{۱۴} توانسته است نقش مهمی در اشتراک‌گذاری ویدئو ایفا و محبوبیت به سزایی در بین کاربران ایرانی پیدا کند. کاربران در این نوع از شبکه‌های اجتماعی، معمولاً ویدئوهایی با مدت زمان کم‌تر از ۵ دقیقه منتشر می‌کنند [۵].

۱-۲-۳- شبکه‌های اجتماعی حرفه‌گرا^{۱۵}

در این نوع شبکه‌های اجتماعی، افراد به منظور پیدا کردن فرصتی مناسب برای بهبود شغل خود، مطالب را به اشتراک می‌گذارند. گرچه برخی از شبکه‌های اجتماعی‌ای که در این گروه قرار می‌گیرند، به طور عمومی در مورد حرفه‌های مختلف صحبت می‌کنند، اما در برخی از موارد، شغل خاصی را مد نظر قرار می‌دهند و در مورد آن به گفت‌وگو می‌نشینند. لینکداین^{۱۶} را می‌توان به عنوان یکی از مهم‌ترین شبکه‌ها در این زمینه اشاره کرد، شبکه‌ای که در سه ماهه‌ی سوم سال ۲۰۱۵، عدد ۳۹۶ میلیون را برای تعداد کاربران خود تجربه کرد^{۱۷}. از دیگر شبکه‌های این بخش می‌توان به کلاس‌روم^{۱۸}، اشاره کرد که محلی است برای معلمان تا بتوانند با یکدیگر در ارتباط باشند.

۱-۲-۴- شبکه‌های اجتماعی انتشار مطالب^{۱۹}

وبلاگ‌ها، بخش اعظمی از این گروه را تشکیل می‌دهند. در این شبکه‌ها، کاربر، مطلب مورد نظر خود را بر روی بستر نرم‌افزاری مورد نظر منتشر می‌کند و افراد پس از مشاهده‌ی مطلب، می‌توانند نظر خود را در مورد آن به کاربر اعلام نمایند. در بسترهای نرم‌افزاری^{۲۰} جدید، امکان به اشتراک‌گذاری و ... نیز برای این مطالب وجود دارد. بلاگفا^{۲۱} را می‌توان از مهم‌ترین سرویس‌دهندگان در این زمینه دانست.

^{۱۳} <https://youtube.com>

^{۱۴} <http://aparat.com>

^{۱۵} Professional Social Networks

^{۱۶} <https://linkedin.com>

^{۱۷} <http://www.statista.com/statistics/274050/quarterly-numbers-of-linkedin-members/>

^{۱۸} <http://classroom20.com>

^{۱۹} Information Publishing Social Networks

^{۲۰} Platforms

^{۲۱} <http://blogfa.com>

۱-۲-۵ - شبکه‌های اجتماعی انجمنی^{۲۲}

از دیگر اقسام شبکه‌های اجتماعی می‌توان به انجمن‌های بحث و تبادل نظر اشاره کرد که در آن‌ها، افراد در مورد یک موضوع یا مطلب خاص با یکدیگر به بحث و گفت‌وگو می‌نشینند. از جمله انجمن‌های معروف می‌توان به دیگ^{۲۳} اشاره کرد. انجمن‌های فارسی‌زبان بسیاری نیز موجود می‌باشند که انجمن تخصصی فناوری اطلاعات ایران^{۲۴} یکی از آن‌ها است.

۱-۲-۶ - شبکه‌های اجتماعی پیام‌رسان^{۲۵}

در این نوع شبکه‌ها، کاربران می‌توانند از طریق ارسال پیام، عکس، ویدئو و صوت، در ارتباط باشند. تلگرام و وایبر^{۲۶} از محبوب‌ترین این شبکه‌ها در ایران می‌باشند. طراحی اختصاصی اپلیکیشن‌های این شبکه‌ها بر روی بستر تلفن همراه و هزینه‌ی بسیار پایین‌تر آن‌ها نسبت به پیامک، باعث رشد روزافزون استفاده از آن‌ها شده است.

۱-۳ - تعریف مسئله و اهمیت آن

با توجه به رشد و نفوذ چشم‌گیر شبکه‌های اجتماعی در بین کاربران بی‌شمار اینترنتی، این شبکه‌ها توانسته‌اند در سال‌های اخیر تغییر و تحولات سیاسی، فرهنگی و اجتماعی شگرفی را در دنیا ایجاد کنند؛ از جمله‌ی این تحولات می‌توان به بهار عربی، و مخصوصاً انقلاب مصر، اشاره کرد [۲، ۳]. از این‌رو، بحث تعیین افراد تاثیرگذار در شبکه‌های اجتماعی با هیجان بیش‌تری از سوی محققان دنبال می‌شود. از آن‌جایی که هر شبکه، ویژگی‌های منحصر به فرد خود را دارد، تشخیص کاربران تاثیرگذار نیز منحصر به آن شبکه می‌شود. از جمله شبکه‌های اجتماعی‌ای که تعریف آن در در بخش قبل رفت، شبکه‌ی اجتماعی انتشار مطلب است. با توجه به ویژگی‌های شبکه‌ی وبلاگ‌ها، اندازه‌گیری میزان تاثیرگذاری نیز نسبت به دیگر شبکه‌ها متفاوت

^{۲۲} Forum Discussion Social Networks

^{۲۳} <http://digg.com>

^{۲۴} <http://itpro.ir>

^{۲۵} Messaging Social Networks

^{۲۶} <http://viber.com/en>

است. از جمله کارهای صورت گرفته در پژوهش حاضر نیز سعی شده است با ارائه‌ی یک مدل جامع در رابطه با شبکه‌ی اجتماعی توئیت، کاربران تأثیرگذار این شبکه معرفی شوند.

نفوذ شبکه‌های اجتماعی در بین آحاد جامعه این مسئله را در پی داشته است که افراد تأثیرگذار نقش بسزایی در جهت‌دهی افکار عمومی خواهند داشت. از این‌رو، جست‌وجو در مورد چنین افرادی در شبکه‌های اجتماعی، مانند توئیت، از اهمیت ویژه‌ای برخوردار است. انتخابات ریاست جمهوری کشورها نیز از دیگر مواردی است که یافتن افراد تأثیرگذار می‌تواند برای نامزدها پر اهمیت باشد؛ زیرا همین افراد هستند که با اظهار نظرات خود می‌توانند ذهن اذعان عمومی را در مورد نامزدهای مختلف تغییر دهند. به طور مثال در [۱]، تحقیقی در رابطه با افرادی که توانستند در شبکه‌ی اجتماعی توئیت، نتایج فعلی انتخابات ریاست جمهوری آمریکا را در سال ۲۰۱۲ رقم بزنند، صورت گرفت.

بازاریابی مجازی از دیگر مواردی است که افراد تأثیرگذار در شبکه‌های اجتماعی می‌توانند در آن نقش کلیدی ایفا کنند. شرکت‌های مختلف در سراسر دنیا، به دنبال ستارگان معروف سینما، تلویزیون و ورزش هستند تا بتوانند محصولات خود را با استفاده از آن‌ها در رسانه‌های عمومی مانند رادیو، تلویزیون و ماهواره به نمایش درآورند. با ورود هنرمندان و ورزشکارها در شبکه‌های اجتماعی پر مخاطب، مانند فیس-بوک، توئیت و اینستاگرام، سبک و سیاق تبلیغات نیز تغییر کرده است و شرکت‌ها تمایل دارند محصولات تولیدی خود را در شبکه‌های مجازی نیز تبلیغ کنند. در چنین شرایطی، نه تنها افراد معروف فضای حقیقی، بلکه افراد کم‌تر شناخته‌شده نیز این فرصت را دارند تا بتوانند از طریق رسانه‌ای جدید، خود را تأثیرگذار جلوه دهند و بازاریاب محصولات شرکت‌های گوناگون باشند.

با توجه به مطالب فوق، می‌توان تشخیص و شناسایی افراد تأثیرگذار در شبکه‌های اجتماعی را یکی از مهم‌ترین مسائل در تحلیل این شبکه‌ها دانست و به آن اشاره کرد؛ لذا در این تحقیق سعی شده با ارائه‌ی یک مدل مناسب، این افراد شناسایی شوند.

۴-۱ - چالش‌ها

در انجام این پژوهش، چالش‌هایی وجود داشتند که هرچند می‌توانستند مانع از انجام و یا ادامه‌ی کار شوند، اما با بهبود یا ارائه‌ی یک روش جایگزین، این چالش‌ها برطرف شدند.

مهم‌ترین چالش انجام پایان‌نامه‌ی حاضر که می‌توانست باعث تغییر پایگاه‌داده شود، فیلتر بودن شبکه‌ی اجتماعی توئیتر بود. این مشکل باعث می‌شد مهم‌ترین اتفاق توئیتر در سال ۲۰۱۵^{۲۷}، در این تحقیق مورد بررسی قرار نگیرد. برای رفع این مشکل، از یک ماشین مجازی بر روی سرور Windows Azure استفاده شد. سرور مورد نظر که مهلت استفاده از آن سه هفته بود، در کشور کانادا قرار داشت. مشکلی که در ادامه‌ی چالش فیلتر بودن توئیتر به وجود آمد، خریداری چنین سروری بود. فراهم نمودن سرور Windows Azure با همکاری یکی از دانش‌جویان ایرانی در خارج از کشور، صورت پذیرفت.

داده‌های استخراج شده در دو مرحله صورت پذیرفتند: در مرحله‌ی اول، اطلاعات مرتبط با توئیتهای و در مرحله‌ی دوم، اطلاعات مرتبط با کاربرانی که توئیتهای مورد نظر را پست کرده‌اند. از آنجایی که مدت استفاده از سرور Windows Azure به سه هفته محدود بود، زمان جمع‌آوری اطلاعات نیز باید در سه هفته به اتمام می‌رسید. با توجه به دو مرحله‌ای بودن جمع‌آوری داده‌ها و نیز حجم بالای آن‌ها، حدود ۴۰ درصد از اطلاعات کاربرانی که توئیتهای را منتشر کرده بودند، به دلیل اتمام مهلت استفاده از سرور مورد نظر، از دسترس خارج شد.

یکی دیگر از چالش‌های موجود در فرآیند پیاده‌سازی مدل ارائه‌شده در این پژوهش، محدودیت رابط کاربردی برنامه‌نویسی‌های توئیتر و گوگل در تعداد درخواست‌ها بود. این محدودیت‌ها باعث کند شدن جمع‌آوری داده و در نتیجه، پیاده‌سازی و نتیجه‌گیری پایان‌نامه شد. برای مثال به منظور استفاده از رابط کاربردی برنامه‌نویسی Google Maps Geocoding، در هر روز، تنها می‌توان ۲۵۰۰ درخواست به آن ارسال کرد. برای ارسال درخواست‌های بیش‌تر می‌بایست به ازای هر ۱۰۰۰ درخواست جدید، مبلغ ۰٫۵ دلار پرداخت شود^{۲۸}. در مورد توئیتر نیز، مثلاً برای استخراج کاربران، تنها می‌توان در هر ۱۵ دقیقه، ۱۸۰ درخواست به رابط کاربردی برنامه‌نویسی توئیتر ارسال کرد^{۲۹}.

^{۲۷} <https://2015.twitter.com/most-influential>

^{۲۸} <https://developers.google.com/maps/documentation/geocoding/usage-limits>

^{۲۹} <https://dev.twitter.com/rest/public/rate-limits>

فصل‌های بعدی پایان‌نامه‌ی حاضر به صورت زیر می‌باشند:

در فصل ۲، کارهای پیشینی که در زمینه‌ی تشخیص افراد تاثیرگذار بوده‌اند، معرفی خواهند شد. فصل ۳ اختصاص دارد به معرفی روش به‌کار گرفته شده در این پژوهش. در فصل ۴ نتایج به دست آمده از اعمال روش ارائه شده، مورد ارزیابی و تحلیل قرار خواهند گرفت و در فصل آخر، نتیجه‌گیری از بحث‌های انجام شده و نیز کارهای آینده خواهند آمد.

۲- مروری بر کارهای پیشین

در فصل جاری به مطالعه و بررسی پژوهش‌های انجام‌شده در زمینه‌ی تشخیص افراد تأثیرگذار در شبکه‌های اجتماعی خواهیم پرداخت. ابتدا معیارهای تأثیرگذاری را از دید نظریه‌ی گراف معرفی می‌کنیم. در بخش بعد، مطالعه‌ی خود را بر روی کارهای انجام‌شده بر روی شبکه‌های اجتماعی وبلاگ‌ها و نیز توئیتر معطوف خواهیم کرد. در قسمت آخر نیز برخی از مهم‌ترین ابزارهای رتبه‌بندی و تحلیلی در زمینه‌ی شبکه‌های اجتماعی معرفی خواهند شد و تحقیقی در این زمینه مورد بررسی قرار خواهد گرفت.

۱-۲- سنجه‌های مرکزیت^{۳۰} در گراف

هر شبکه‌ی اجتماعی در زبان ریاضی می‌تواند به یک گراف مدل شود. از این‌رو می‌توان سنجه‌های مرکزیت در یک گراف را برای هر گره^{۳۱} محاسبه نمود و از آن جایی که هر گره در گراف، معادل یک کاربر در شبکه‌ی اجتماعی است، می‌توان هر یک از سنجه‌های مرکزیت را به عنوان یکی از عوامل تأثیرگذاری فرد در شبکه‌ی اجتماعی در نظر گرفت. سنجه‌های مرکزیت به شکل زیر تعریف می‌شوند:

۱-۱-۲- لینک ورودی^{۳۲}

این سنجه در واقع همان لینک‌های ورودی به یک گره در گراف است. در یک شبکه‌ی اجتماعی، این عامل را می‌توان به عنوان دوستانی دانست که با یک کاربر در ارتباط هستند. هر چه تعداد دوستان یک فرد بیشتر باشد، میزان تأثیرگذاری فرد مورد نظر در شبکه‌ی اجتماعی بیشتر است.

۲-۱-۲- بینابینی^{۳۳}

در نظریه‌ی گراف^{۳۴}، اگر گره‌ای مانند A داشته باشیم که اکثر کوتاه‌ترین مسیرهای بین هر دو گره از A بگذرد، این گره دارای بینابینی بالایی است [۶]. رابطه‌ی (۱-۲) تعریف فوق را به صورت ریاضی نشان می‌دهد که در آن g_{jk} بیان‌گر تعداد کل کوتاه‌ترین مسیرهایی است که از گره‌های j و k می‌گذرد و $g_{jk}(i)$ تعداد کوتاه‌ترین مسیرهای بین j و k است که از گره‌ی i می‌گذرد. مثالی بصری از سنجه‌ی بینابینی در شکل ۱-۲ آمده است. همان‌طور که در این

^{۳۰} Centrality measures

^{۳۱} Node

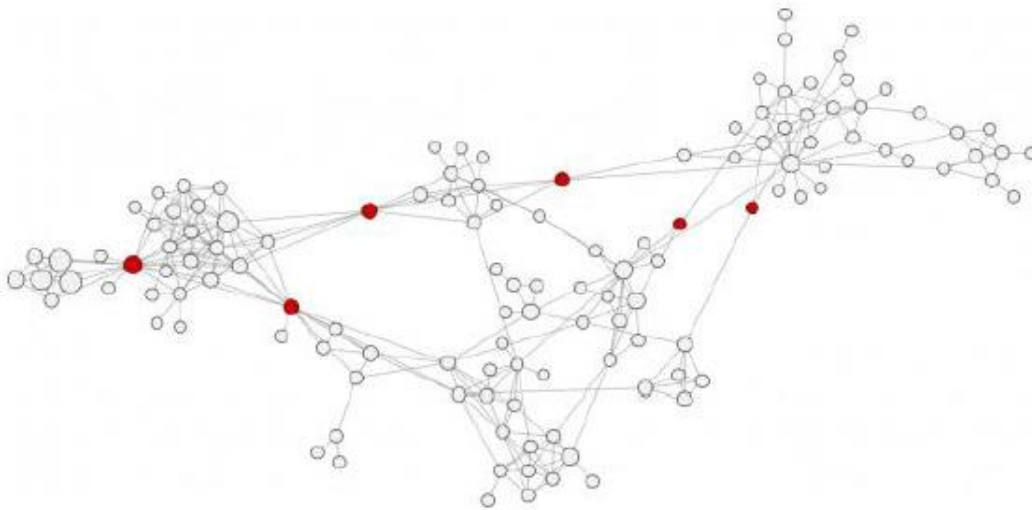
^{۳۲} Inlink

^{۳۳} Betweenness

^{۳۴} Graph theory

شکل مشخص است، گره‌هایی دارای بینابینی بالایی هستند که غالباً ارتباط-
 دهنده‌ی بین دو یا چند اجتماع^{۳۵} می‌باشند؛ لذا این افراد تأثیر به‌سزایی در
 رابطه‌ی بین دو اجتماع دارند.

$$B(i) = \sum_{j < k} g_{jk}(i) / g_{jk} \quad (1-2)$$



شکل ۱-۲ مثالی بصری از سنجه‌ی بینابینی^{۳۶}

۳-۱-۲- نزدیکی^{۳۷}

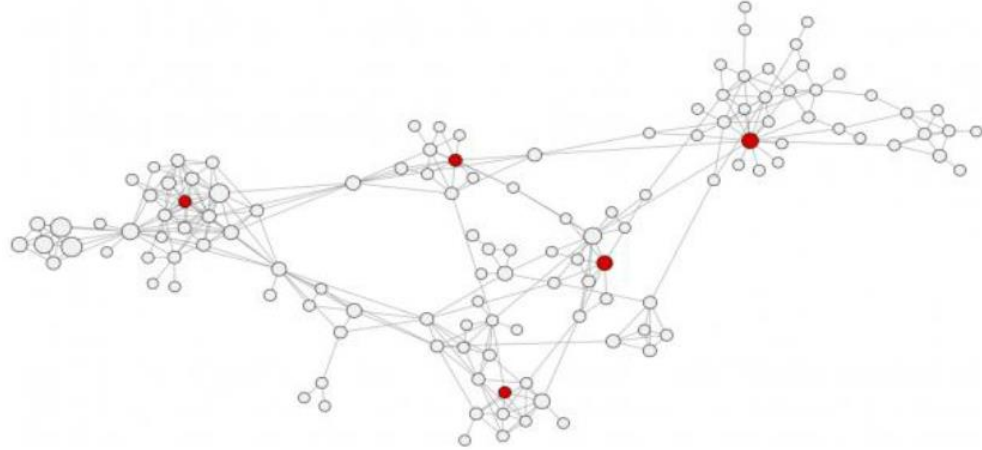
این سنجه بیان‌گر متوسط فاصله‌ی بین هر گره تا گره‌های دیگر است. در یک
 شبکه‌ی اجتماعی، افرادی که با کوتاه‌ترین مسیر به بیش‌ترین افراد متصل می-
 شوند، دارای نزدیکی بالایی می‌باشند. مدل ریاضی این سنجه در رابطه‌ی (۲-۲)
 آمده است که در آن، $d(i,j)$ فاصله‌ی بین گره‌ی i با دیگر گره‌ها مانند j است.
 مثالی از مدل بصری این سنجه نیز در شکل ۲-۲ آمده است. همان‌طور که
 ملاحظه می‌شود، افراد با نزدیکی بالا از آن جهت تأثیرگذار هستند که با کم‌ترین
 فاصله‌ی ممکن می‌توان از آن‌ها به افراد بسیاری در شبکه رسید.

^{۳۵} Community

^{۳۶} برگرفته شده از <https://www.linkedin.com/pulse/wtf-do-you-actually-know-who-influencers-walter-pike>

^{۳۷} Closeness

$$C(i) = [\sum_{j=1}^N d(i,j)]^{-1} \quad (2-2)$$



شکل ۲-۲ مثالی بصری از سنجهی نزدیکی^{۴۰}

۲-۲- مدلهای خاص

۲-۲-۱- وبلاگها

در تحقیق انجام شده در [۷] که در سال ۲۰۰۸ انجام شد، نیتین^{۳۸} و همکارانش به بررسی تأثیرگذاری وبلاگها در سرویس فراهم کنندهی وبلاگی^{۳۹} Unofficial Apple Weblog (TUAW) پرداختند. مهم ترین مسئلهای که در تحقیق مورد نظر به آن اشاره شده است، تفاوت بین وبلاگ نویسان تأثیرگذار^{۴۰} و فعال^{۴۱} است. به عبارت دیگر، هر وبلاگ نویسی که تعداد پستهای زیادی منتشر کند، الزاماً

^{۳۸} Nitin Agarwal

^{۳۹} Weblog service providers

^{۴۰} Influential

^{۴۱} Active

نمی‌تواند وبلاگ نویس تأثیرگذاری باشد. در واقع، تعداد پست‌ها هیچ رابطه‌ای با میزان تأثیرگذاری پست‌ها ندارد. همان‌طور که در این تحقیق اشاره شده است، برای این که وبلاگی تأثیرگذار شناخته شود، باید ۴ ویژگی زیر را داشته باشد:

- **شناخت^{۴۲}**: منظور از شناخت این است که یک وبلاگ تأثیرگذار باید توسط بسیاری از وبلاگ‌های دیگر شناخته شده باشد. برای این منظور، پستی مانند p زمانی می‌تواند تأثیرگذار باشد که توسط بسیاری از پست‌های دیگر ارجاع داده شده باشد یا در واقع تعداد لینک‌های ورودی آن به اندازه‌ی l بزرگ باشد.
- **میزان فعالیت^{۴۳}**: میزان فعالیت یک پست را می‌توان به طور غیرمستقیم از تعداد نظرات آن پست اندازه‌گیری کرد. در واقع هنگامی که پستی به تعداد γ نظر دریافت می‌کند، بدین معنی است که این پست توانسته است بر روی افراد زیادی تأثیر بگذارد. یکی از مهم‌ترین دغدغه‌ها در مورد این ویژگی، نظراتی است که توسط هرزنامه‌نویسان^{۴۴} تولید می‌شود [۷].
- **خلاقیت^{۴۵}**: تعداد لینک‌های خروجی^{۴۶} هر پست، بیان‌گر میزان خلاقیت آن است. هر چه تعداد لینک‌های خروجی به عدد θ نزدیک‌تر باشد، میزان خلاقیت آن کم‌تر است. با توجه به مطلب گفته‌شده، می‌توان این‌گونه بیان کرد که تعداد لینک‌های خروجی رابطه‌ی معکوس با تعداد نظرات دارد.
- **فصاحت^{۴۷}**: اندازه‌گیری این ویژگی با استفاده از ابزارهای آماری، در مقایسه با ویژگی‌های دیگر، سخت‌تر است. از آنجایی که وبلاگ‌ها غالباً به گونه‌ای هستند که به مخاطب خود اطلاعاتی را ارائه می‌کنند، تعداد کاراکترهای^{۴۸} بکار گرفته شده در هر پست، می‌تواند بیان‌گر تأثیرگذاری آن پست باشد. به عبارت دیگر، هر چه تعداد کاراکترهای پست به عدد λ نزدیک‌تر باشد، تأثیرگذاری آن بیش‌تر است.

^{۴۲} Recognition

^{۴۳} Activity generation

^{۴۴} Spammers

^{۴۵} Novelty

^{۴۶} Outlink

^{۴۷} Eloquence

^{۴۸} Characters

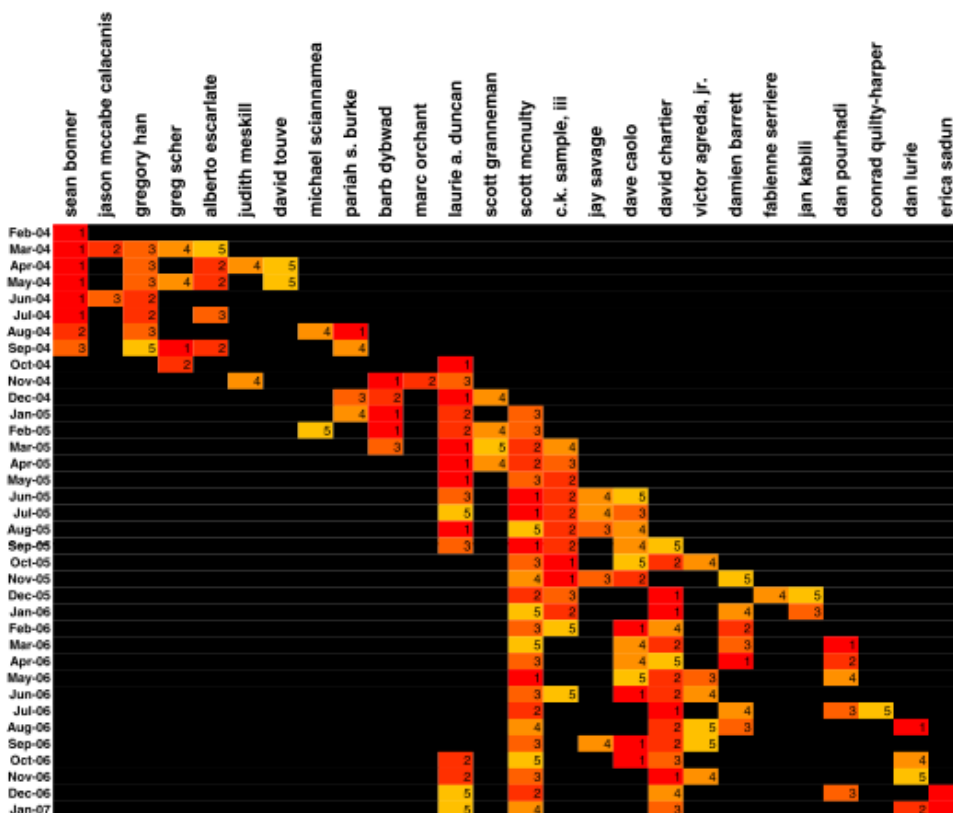
در مدل ارائه شده، که **Influence Flow** نام دارد، هر ۴ ویژگی ذکر شده مورد استفاده قرار می-گیرند. برای محاسبه‌ی میزان تأثیرگذاری هر پست، از رابطه‌ی (۳-۲) استفاده می‌شود. در فرمول ارائه شده، برای لینک‌های ورودی، خروجی و تعداد کاراکترهای هر پست، وزنی در بازه‌ی [۰-۱] در نظر گرفته می‌شود.

$$I(p) = w(\lambda) * (w_{com} \gamma_p + w_{in} \sum_{m=1}^{|I|} I(p_m) - w_{out} \sum_{n=1}^{|\theta|} I(p_n)) \quad (3-2)$$

در این تحقیق، بیش از ۱۰۰۰۰ وبلاگ مورد بررسی قرار گرفتند. پس از محاسبه و استخراج تأثیرگذارترین وبلاگ‌ها، نتایج به صورت زیر مورد تحلیل قرار گرفتند:

- **وبلاگ‌نویسان فعال و تأثیرگذار:** در این گروه، وبلاگ‌نویسانی قرار می‌گیرند که هم تعداد پست‌های زیادی داشته‌اند و هم این‌که پست‌های آن‌ها با تعداد نظرات و لینک‌های ورودی زیادی همراه بوده است.
- **وبلاگ‌نویسان غیرفعال اما تأثیرگذار:** این گروه از وبلاگ‌نویسان، آن‌هایی هستند که بر خلاف تعداد پست‌های کمی که داشته‌اند، توانسته‌اند مخاطب بسیاری را به خود جلب کنند و هر یک از معدود پست‌های آن‌ها دارای نظرات و لینک‌های ورودی بسیاری زیادی باشند. یکی از وبلاگ‌نویسانی که در این گروه وجود دارد، پستی در رابطه با iPhone منتشر کرده است که ۷۷ نظر و ۳۳ لینک ورودی را به خود اختصاص داد. با مراجعه به متن اصلی این پست می‌توان درک کرد که وبلاگ‌نویس تجربه‌ی شخصی و جزئی خود را در مورد این محصول به مخاطبان ارائه داده است.
- **وبلاگ‌نویسان فعال اما غیرتأثیرگذار:** این وبلاگ‌نویسان را افرادی تشکیل می‌دهند که تعداد پست‌های بالایی را منتشر می‌کنند، اما محتویات پست‌هایشان برای مخاطبان، جاذبه‌ی خاصی ایجاد نمی‌کند.

تحقیق مورد نظر در یک بازه‌ی ۳ ساله صورت گرفته است که افراد تأثیرگذار در آن به صورت بصری در شکل ۳-۲ آمده است.



شکل ۲-۳ نمودار افراد تأثیرگذار در یک بازه‌ی ۳ ساله با توجه به مدل [۲۴] *Influence Flow*

با توجه به شکل ۲-۳، افراد تأثیرگذار را می‌توان در ۴ گروه طبقه‌بندی کرد:

- **افراد تأثیرگذار بلندمدت:** این افراد، توانسته‌اند تأثیرگذاری خود را به طور منظم حفظ کنند. به عبارت دیگر، این وبلاگ‌نویسان بر فضای وبلاگ‌نویسی تسلط دارند. Scott McNulty بارزترین مثال در شکل ۲-۳ است.
- **افراد تأثیرگذار میان مدت:** در این گروه، افرادی قرار می‌گیرند که توانسته‌اند تأثیرگذاری خود را برای ۴-۵ ماه حفظ کنند. با توجه به شکل ۲-۳، این افراد شامل Sean Bonner، Gregory Han و Barb Dybward می‌شوند.

• افراد تأثیرگذار کوتاه مدت: این افراد برای یک مدت بسیار کوتاه (۱-۲ ماه) توانسته‌اند در لیست وبلاگ‌نویسان تأثیرگذار قرار بگیرند. Michael Sciannamea و Fabienne Serriere از جمله مثال‌های شکل ۲-۳ می‌باشند.

• افراد تأثیرگذار در حال رشد: وبلاگ‌نویسانی که در این گروه قرار می‌گیرند، به تازگی به عنوان افراد تأثیرگذار در شبکه‌ی وبلاگ‌ها شناخته شده‌اند و باید منتظر بود تا دید در آینده چه بر سر آن‌ها خواهد آمد. با نگاه به شکل ۲-۳، Dan Lurie و Erica Sadun از جمله‌ی این افراد می‌باشند.

مهم‌ترین عاملی که در [۷] از آن چشم‌پوشی شده، زمان است. در واقع، مدل Influence Flow به گونه‌ای ارائه شده است که اگر وبلاگ‌نویسی از زمان t_0 تا t_1 پست، نظر، لینک ورودی و خروجی نداشته باشد، میزان تأثیرگذاری وی نیز تغییر نمی‌کند. این مشکل اساسی باعث شد تا یک سال بعد (سال ۲۰۰۹) لئونیداس^{۴۹} و همکارانش مدل ارائه شده در [۷] را اصلاح نموده و با اضافه نمودن عامل زمان، کار خود را در [۸] ارائه نمایند. مدل ارائه شده MEIBI^{۵۰} نام دارد. در این مدل، اختلاف زمان بین زمان حال و زمانی که پستی در وبلاگ منتشر شده است، با میزان تأثیرگذاری ارتباط معکوس دارد؛ به عبارت دیگر، هر چه این اختلاف زمانی بیشتر باشد، میزان تأثیرگذاری کاهش می‌یابد. رابطه‌ی (۲-۴) این مدل را به صورت ریاضی تشریح می‌کند که در آن، $S_j^m(i)$ میزان تأثیرگذاری i -امین پست وبلاگ‌نویس j اندازه‌گیری می‌شود.

$$S_j^m(i) = \gamma(|C(i) + 1| * (\Delta TP_j(i) + 1)^{-\delta} |R_j(i)|) \quad (۲-۴)$$

پارامترهای مدل به صورت زیر می‌باشند:

- $C(i)$: نظر(هایی) که برای پست i داده می‌شود؛
- $R_j(i)$: لینک‌های ورودی به پست i ؛
- $\Delta TP_j(i)$: اختلاف زمانی بین زمان حال و زمانی که پست i منتشر شده است؛
- γ و δ : ضرایب انتخابی برای بزرگ‌تر کردن اندازه‌ی تأثیرگذاری.

^{۴۹} Leonidas Akritidis

^{۵۰} Metric for Evaluating and Identifying a Blogger's Influence

ایرادی که به مدل MEIBI وارد می‌شود این است که اگر یک پست پس از مدت‌های طولانی همچنان لینک ورودی بپذیرد یا به عبارتی اگر پست‌هایی که اخیراً منتشر می‌شوند، مطلب خود را به یک پست قدیمی ارجاع دهند، میزان تأثیرگذاری پست قدیمی بدون تغییر باقی می‌ماند؛ لذا در [۸]، مدل دیگری به نام MEIBIX^{۵۱} معرفی می‌شود که بتواند این نقص را نیز برطرف سازد. مدل MEIBIX به صورت (۲-۵) فرمول‌سازی می‌شود:

$$S_j^x(i) = \gamma(|C(i) + 1| * \sum_{v \in R_j(i)} (\Delta TP_j(i) + 1))^{-\delta} \quad (۵-۲)$$

برای اندازه‌گیری میزان تأثیرگذاری وبلاگ‌نویسان، از پست‌های ارسال شده در سرویس وبلاگ TUAW استفاده شده است. برای این کار، ۱۷۸۳۱ پست که توسط ۵۱ وبلاگ‌نویس منتشر شده‌اند، جمع‌آوری شده است. تعداد کل نظرات برای این تعداد پست نیز عدد ۲۶۹۴۴۹ گزارش شده است. از کل پست‌های جمع‌آوری شده، تنها ۴۵۸۶ پست دارای لینک ورودی می‌باشند. شکل ۲-۱ آماری از تعداد پست‌ها و لینک‌های ورودی را در این پایگاه داده بیان می‌کند.

جدول ۲-۱ آمار پست‌ها و لینک‌های ورودی پایگاه داده‌ی مورد استفاده در [۸]

سال	تعداد پست‌ها	تعداد پست‌های دارای لینک ورودی	تعداد لینک‌های ورودی
۲۰۰۴	۹۹۷	۸	۰
۲۰۰۵	۴۳۰۷	۷۷	۱
۲۰۰۶	۴۳۵۴	۱۸۶	۱۸
۲۰۰۷	۴۴۹۷	۶۶۲	۲۵۹
۲۰۰۸	۳۶۷۶	۳۶۵۳	۵۳۲۰۴
کل	۱۷۸۳۱	۴۵۸۶	۵۳۵۷۵

با توجه به محاسبات انجام‌شده، نتایج حاصل برای مدل‌های MEIBI، Influence Flow و MEIBIX در جدول ۲-۲ آمده است. برای اندازه‌گیری میزان همبستگی بین مدل‌های مذکور از Spearman's rho استفاده شده است. جدول ۲-۳ رابطه‌ی بین این ۳ روش را نشان می‌دهد. نتایج به دست آمده حاکی از عدم وابستگی بین مدل ارائه شده در [۷] و مدل‌های معرفی شده در [۸] است.

^{۵۱} MEIBI eXtended

جدول ۲-۲ نتایج بدست آمده از روش‌های *MEIBI Influence Flow* و *MEIBIX*؛ راست: *Influence Flow*؛ وسط: *MEIBI*؛ چپ: *MEIBIX* [۷]

x	وبلاگ‌نویس		m	وبلاگ‌نویس		وبلاگ‌نویس	
27	C. Bohon	۱	26	C. Bohon	۱	C. Bohon	۱
20	S. Sande	۲	20	R. Palmer	۲	R. Palmer	۲
19	R. Palmer	۳	20	S. Sande	۳	M. Lu	۳
18	D. Caolo	۴	17	D. Caolo	۴	C. Warren	۴
16	M. Schramm	۵	16	M. Schramm	۵	D. Caolo	۵
13	M. Rose	۶	13	M. Rose	۶	C. Ullrich	۶
8	M. Lu	۷	8	M. Lu	۷	S. Sande	۷
7	B. Terpstra	۸	7	B. Terpstra	۸	M. Rose	۸
7	C. Warren	۹	7	C. Warren	۹	V. Agreda	۹
4	V. Agreda	۱۰	4	V. Agreda	۱۰	Jason Clarke	۱۰

جدول ۳-۲ میزان همبستگی بین مدل‌های *MEIBI Influence Flow* و *MEIBIX* [۸]

ρ	روش‌ها
۰,۴۱۸۱۸۲	Influence Flow – MEIBI
۰,۳۵۷۵۷۶	Influence Flow – MEIBIX
۰,۹۸۷۸۷۹	MEIBI – MEIBIX

در ادامه‌ی پژوهش [۸]، لئونیداس و همکارانش در سال ۲۰۱۱ در [۹] نیز به تعیین افراد تأثیرگذار می‌پردازند و در قسمتی از پژوهش، خود را معطوف به معرفی مدلی برای تشخیص افراد مولد^{۵۲} در وبلاگ‌ها می‌کنند. آن‌ها ابتدا به معرفی انواع تأثیرگذاری می‌پردازند و آن را به دو نوع زیر تقسیم می‌کنند:

^{۵۲} Productive

۱. تأثیر محدود^{۵۳}: این نوع اثرگذاری زمانی است که یک وبلاگ‌نویس بر روی اعضای می‌تواند تأثیر بگذارد که در اجتماع^{۵۴} وی حضور دارند. این نوع اثر را غالباً می‌توان با تعداد نظرات یک پست مشاهده نمود.

۲. تأثیر گسترده^{۵۵}: در این نوع تأثیرگذاری، وبلاگ‌نویس می‌تواند بر روی افرادی که در اجتماع وی نیستند نیز اثر بگذارد. تعداد لینک‌های ورودی هر پست، بیان‌گر این نوع موضوع است. اگرچه برخی ویژگی‌های دیگری مانند تعداد لایک‌ها^{۵۶} در فیس‌بوک نیز می‌تواند این نوع تأثیر را نشان دهد، اما سرویس‌های وبلاگی محدودی دارای این ویژگی هستند.

برای تعیین افراد مولد در فضای وبلاگ‌ها، رابطه‌ی (۲-۶) در [۹] ارائه شده است:

$$U_{i,p}^j(t) = \gamma \frac{L_i^j}{\bar{L}} \left(\frac{\theta}{t - t_{i,p}^j + \theta} \right)^\delta \quad (۲-۶)$$

که در آن، پارامترهای به کار گرفته شده به صورت زیر تعریف می‌شوند:

- L_i^j : اندازه (تعداد کاراکترهای) i -امین پست وبلاگ‌نویس j ؛
- \bar{L} : متوسط اندازه‌ی پست؛
- $t_{i,p}^j$: زمان ارسال پست i از وبلاگ‌نویس j ؛
- γ, θ, δ : پارامترهای ثابت مدل با مقادیر دلخواه.

مدل BP^{۵۷} برای تشخیص وبلاگ‌نویسان مولد بیان می‌کند که در زمان t ، مقدار شاخص BP برای وبلاگ‌نویس j برابر P_i^j است، اگر و تنها اگر P_i^j تا از پست‌های وبلاگ‌نویس، هر کدام، دارای مقدار $U_{i,p}^j(t) \geq P_i^j$ باشند.

برای تعیین افراد تأثیرگذار نیز، مدل BI^{۵۸} معرفی شده است. اگر

^{۵۳} Proximal impact

^{۵۴} Community

^{۵۵} Wide impact

^{۵۶} Like

^{۵۷} Blogger's Productivity

^{۵۸} Blogger's Influence

$$V_{i,p}^j(t) = w_l \sum_{v_x \in R_i^j} \left(\frac{\theta}{t - t_{x,l}^j + \theta} \right)^\delta + w_c \sum_{v_x \in C_i^j} \left(\frac{\theta}{t - t_{x,c}^j + \theta} \right)^\delta \quad (7-2)$$

را با پارامترهای

- R_i^j : لینک‌های ورودی پست i ؛
- C_i^j : نظرات پست i

داشته باشیم، آنگاه مدل BI بیان می‌کند در زمان t ، مقدار شاخص BI برای وبلاگ نویس j برابر P_i^j است، اگر و تنها اگر P_i^j تا از پست‌های وبلاگ‌نویس، هر کدام، دارای مقدار $V_{i,p}^j(t) \geq P_i^j$ باشند.

برای انجام آزمایش‌ها و نتیجه‌گیری، پایگاه داده‌ای از وبلاگ‌های موجود در سرویس Engadget^{۵۹} جمع‌آوری و مدل‌های ارائه شده، در آن مورد بررسی قرار گرفتند. اطلاعات پایگاه‌داده‌ی مورد نظر در جدول ۲-۴ آمده است.

جدول ۲-۴ اطلاعات پایگاه‌داده‌ی استفاده‌شده در [۹]

عنوان	تعداد
وبلاگ‌نویسان	۹۳
پست‌ها	۶۳۳۵۸
لینک‌های ورودی	۳۱۹۸۸۰
نظرات	۳۶۷۲۸۱۹
میانگین طول هر پست (کلمه)	۱۸۰٫۵۲

برای ارزیابی مدل‌های BP و BI، این دو، با مدل‌های MEIBI و MEIBIX مقایسه شده‌اند. مقایسه‌ی مدل‌ها در جدول ۲-۵ آمده است.

جدول ۲-۵ مقایسه‌ی مدل‌های BP، BI، MEIBI و MEIBIX [۹]

ردیف	وبلاگ‌نویس	BP	BI	MEIBI	MEIBIX
۱	D. Murph	۳۲	۱۱۴	۱۱۳	۱۳۴

^{۵۹} <http://engadget.com>

۱۲۹	۱۱۳	۹۴	۲۶	C. Ziegler	۲
۱۲۲	۱۰۱	۹۰	۲۶	V. Savov	۳
۱۰۵	۸۹	۶۸	۲۵	P. Miller	۴
۶۷	۵۵	۵۸	۲۴	J. L. Flatley	۵
۶۸	۶۰	۶۴	۲۱	T. Stevens	۶
۵۰	۴۵	۴۱	۲۱	J. Stern	۷
۱۰۷	۹۲	۷۹	۲۰	T. Ricker	۸
۱۲۵	۱۱۳	۷۲	۲۰	J. Topolsky	۹
۸۵	۷۷	۵۵	۲۰	R. Miller	۱۰

نتایج به دست آمده نشان می‌دهد که رتبه‌بندی طبق مدل‌های MEIBI و MEIBIX با آن چه در BP و BI به دست آمده، متفاوت است و تطابقی ندارد. این نتایج از آن جهت است که مدل MEIBI، تأثیری که اخیراً وبلاگ‌نویسان داشته‌اند^{۶۰} را در نظر نمی‌گیرد. همین مطلب برای MEIBIX صادق است؛ زیرا این مدل نیز مولد بودن اخیر^{۶۱} یک وبلاگ‌نویس را در خود اثر نمی‌دهد. از طرف دیگر، این دو مدل، اختلاف زمانی (سن) نظرات یک پست را نیز در نظر نمی‌گیرند.

با توجه به روند رو به جلوی سرویس‌های ارائه‌دهنده‌ی وبلاگ و امکاناتی که برای سهولت ارتباط وبلاگ‌نویسان با دیگر شبکه‌های اجتماعی فراهم می‌کنند، در برخی از سرویس‌ها می‌توان امکاناتی مانند به اشتراک‌گذاری پست در شبکه‌های اجتماعی دیگر مانند فیس‌بوک و نیز امکان لایک کردن پست را مشاهده کرد. با توجه به این امکانات جدید، در سال ۲۰۱۳، تنگ‌شن^{۶۲} از دانشگاه ایالتی سن‌خوزه^{۶۳} در [۱۰] مدل پیشنهادی خود را بر روی وبلاگ‌ها ارائه داد. در این مدل برای تشخیص تأثیرگذاری، ۵ عامل معرفی شده است که به صورت زیر شرح داده می‌شوند:

- امتیاز فیس‌بوک و لینک‌های ورودی^{۶۴}: این عامل، غیر از در نظر گرفتن اهمیت لینک-های ورودی در میزان تأثیرگذاری، تعداد لایک‌ها و به اشتراک‌گذاری‌های پست‌ها در فیس-بوک را نیز از عوامل مهم در نقش وبلاگ‌نویسان برمی‌شمارد.

^{۶۰} Recent influence

^{۶۱} Recent productivity

^{۶۲} Teng-Shen Moh

^{۶۳} San Jose State University

^{۶۴} FB score and inlink score

- **عامل حسی^{۶۵}**: از آنجایی که مخاطبان وبلاگ با خواندن هر پست، نظرات خود را در مورد آن می‌نویسند، می‌توان احساس مخاطبان نسبت به هر پست را، مثبت، منفی و یا خنثی در نظر گرفت. مثبت و یا منفی بودن نظرات، از دیگر عوامل موثر در تأثیرگذاری یک پست، و در نتیجه یک وبلاگ‌نویس، است.
- **تأثیر درج‌کننده‌ی نظر در هر پست^{۶۶}**: از آنجایی که درج‌کننده‌ی نظر می‌تواند یک کاربر عادی و یا یک متخصص باشد، میزان تأثیر نظر وی در هر پست متفاوت است. مثلاً اگر یک متخصص در همان موضوعی که پست گذاشته شده است در مورد آن پست نظر دهد، این نظر با تأثیر بیش‌تری نسبت به نظر یک کاربر عادی همراه خواهد بود.
- **عامل زمان: زمان**، به عنوان یکی از مهم‌ترین عوامل، می‌تواند در تأثیرگذاری افراد موثر باشد؛ لذا با افزایش زمان، تأثیرگذاری یک پست کم خواهد شد.
- **یکتایی پست و لینک‌های خروجی^{۶۷}**: این عامل نیز، مانند زمان، رابطه‌ی معکوس با میزان تأثیرگذاری دارد. در واقع هرچه تعداد لینک‌های موجود در پست زیاد باشد، آن پست از خلاقیت نویسنده‌ی آن می‌کاهد و یکتایی خود را از دست می‌دهد و لذا تأثیر کم‌تری دارد.

با توجه به موارد ذکرشده، هر یک از این عوامل به صورت ریاضی مدل می‌شوند. هر یک از این مدل‌ها در زیر آمده است:

- **امتیاز فیس‌بوک**: این امتیاز به صورت (۸-۲) محاسبه می‌شود که در آن، Fb_i^j جمع تعداد لایک‌ها و دفعات به اشتراک‌گذاری شدن پست است.

$$Fb_{i,p}^j(t) = w_{fb} * Fb_i^j * \left(\frac{\theta}{t-t_{x,p}^j+\theta}\right)^\delta \quad (8-2)$$

- **مقدار تأثیر لینک‌های ورودی**: این مقدار با استفاده از (۹-۲) قابل اندازه‌گیری است.

$$INLK_{i,p}^j(t) = w_l \sum_{v_x \in R_i^j} \left(\frac{\theta}{t-t_{x,l}^j+\theta}\right)^\delta \quad (9-2)$$

^{۶۵} Sentimental factor

^{۶۶} Influence of commenters

^{۶۷} Uniqueness and outlinks

- **عامل حسی و نقش درج‌کننده‌ی نظر:** برای این بخش، اگر SF بیان‌گر مثبت، منفی و یا خنثی بودن نظر باشد، Inf همان سطح تأثیرگذاری درج‌کننده‌ی نظر باشد و $TC(b_c)$ تعداد کل نظرهای داده‌شده برای پست مورد نظر باشد، اندازه‌ی تأثیر نظرات پست با (۱۰-۲) قابل محاسبه است.

$$C_{i,p}^j(t) = w_c \sum_{v_x \in C_i^j} \left(\frac{Inf(b_c) * SF(b_c, x)}{TC(b_c)} \right) * \left(\frac{\theta}{t - t_{x,c}^j + \theta} \right)^\delta \quad (10-2)$$

- **خلاقیت:** برای این فاکتور، ابتدا کلمات پر تکرار^{۶۸} هر پست حذف شده‌اند. سپس با استفاده از الگوریتم TFIDF کلمات مهم پست استخراج شده‌اند. برای محاسبه‌ی میزان خلاقیت پست، از رابطه‌ی (۱۱-۲) کمک گرفته شده است.

$$N_i^j = w_n \left(1 - \left(\frac{\text{number of words matching the key words}}{\text{total number of words}} \right) - w_o * \text{outlinks} \right) \quad (11-2)$$

برای محاسبه‌ی میزان تأثیرگذاری هر پست نیز از رابطه‌ی (۱۲-۲) استفاده شده است:

$$PS_{i,p}^j = Fb_{i,p}^j + C_{i,p}^j + INLK_{i,p}^j + N_i^j \quad (12-2)$$

پس از انجام عمل مدل‌سازی، این مدل در پایگاه‌داده‌ی ارائه شده در [۹]، به کار گرفته شده

است.

۲-۲-۲- توئیتر

توئیتر به عنوان یکی از شبکه‌های اجتماعی موفق در بین کاربران اینترنتی، کار خود را از ۲۰۰۶ و با همکاری جک دورسی^{۶۹}، نوآ گلس^{۷۰}، ایوان ویلیامز^{۷۱} و بیز استون^{۷۲} به صورت رسمی آغاز کرد. هر ماه، نزدیک به ۳۲۰ میلیون کاربر فعال در این شبکه‌ی محبوب اجتماعی فعالیت می‌کنند^{۷۳}. محبوبیت این پرنده‌ی خوش‌آواز باعث شده است پژوهش‌گران مختلف در

^{۶۸} Stop words

^{۶۹} Jack Dorsey

^{۷۰} Noah Glass

^{۷۱} Evan Williams

^{۷۲} Biz Stone

^{۷۳} <https://about.twitter.com/company>

زمینه‌های علوم اجتماعی، سیاست، اقتصاد، کامپیوتر، ریاضی و ... تمرکز خود را بر روی آن بگذارند تا بتوانند علل برخی از تغییر و تحولات جامعه و جهان را در این شبکه جست‌وجو کنند [۲، ۳، ۱۱، ۱۲، ۱۳]. در این بخش به برخی از پژوهش‌های صورت گرفته در شبکه‌ی اجتماعی توئیتر که در رابطه با تشخیص افراد تأثیرگذار است، می‌پردازیم. قبل از شروع به بررسی پژوهش‌های انجام شده در توئیتر، بهتر است برای آشنایی با برخی از مفاهیم این شبکه‌ی اجتماعی، به بخش ۳-۱ مراجعه شود.

بهار عربی یکی از مهم‌ترین اتفاقات سال‌های اخیر در زمینه‌ی سیاسی است که بازتاب گسترده‌ای در شبکه‌های اجتماعی و مخصوصاً توئیتر داشته است. از جمله پژوهش‌های انجام شده در این زمینه، کاری است که لوکاس و همکارانش در سال ۲۰۱۳ در [۲] منتشر کردند. در این تحقیق، الگوریتم آلفا [۱۴] برای تشخیص مهم‌ترین افراد در توئیتر به‌کاربرده شده است. برای تشخیص تأثیرگذارترین افراد در انقلاب سال ۲۰۱۱ در مصر، توئیتهای مربوط به کاربران در بازه‌ی ۱۰ دسامبر ۲۰۱۰ تا ۱۱ اکتبر ۲۰۱۱ که در آن از هشتگ‌های^{۲۴} #Jan25، #Egypt، #mubarak و #tahrir استفاده شده است، جمع‌آوری شدند. پس از جمع‌آوری داده‌ی مورد نظر و اعمال مدل آلفا، مشخص شد تأثیرگذارترین کاربران توئیتر در هر هشتگ، عموماً خبرگزاری‌هایی هستند که تغییر و تحولات انقلاب مصر را پوشش خبری داده‌اند. جدول ۲-۶ نتایج این پژوهش را مشخص می‌نماید.

جدول ۲-۶ نتایج به دست آمده از تأثیرگذارترین افراد پس از اعمال مدل آلفا [۲]

هشتگ				رتبه
#mubarak	#tahrir	#Jan25	#Egypt	
shorouk news	Lilianwagdy	Dostornews	dostornews	۱
Monaeltahawy	gsquare86	Ghonim	Earhim	۲
Adamakary	Carloslatuff	Sandmonkey	Ghonim	۳
Lilianwagdy	shorouk news	Alaa	almasryalyoum a	۴
Raafatology	Leilzahra	arabawy۳	Ontveg	۵
almasryalyoum a	Salmasaid	almasryalyoum a	arabawy۳	۶

^{۲۴} برای توضیح در مورد هشتگ به بخش ۳-۱ مراجعه فرمایید

Egyptocracy	arabawy۳	Monasosh	Egyfeeds	۷
Carloslatuff	Pakinamamer	dima khatib	Alaa	۸
almasryalyoum e	Monasosh	Wael	Carloslatuff	۹
Daliaziada	Sandmonkey	Egyptocracy	Zeinobia	۱۰

در تحقیق دیگری که در رابطه با بهار عربی در سال ۲۰۱۳ انجام شد، می‌توان به پژوهش کومار و همکارانش پرداخت که در [۳] به بررسی افراد تأثیرگذاری پرداخت که در هنگام مواجهه با بحران، می‌توان آن‌ها را شکل‌دهنده‌ی افکار مردم دانست. در این تحقیق، توئیت‌هایی جمع‌آوری شده‌اند که نویسنده‌ی آن یا از کشورهای مصر، لیبی، سوریه، تونس و یمن پیام خود را ارسال کرده و یا در توئیت‌های ارسالی، نام یکی از کشورهای مذکور بوده است. دو رکن اصلی این تحقیق برای تشخیص افراد تأثیرگذار عبارت‌اند از: (۱) مرتبط بودن محتوای توئیت با بهار عربی؛ (۲) ارسال شدن توئیت از یکی از کشورهای نام‌برده. برای تشخیص مرتبط بودن محتوای هر توئیت با موضوع انقلاب‌های صورت گرفته در کشورهای مصر، لیبی، سوریه، تونس و یمن، ابتدا از الگوریتم LDA استفاده شده است. پس از تعیین موضوع هر توئیت، توئیت‌هایی که موضوع‌شان با بهار عربی هم‌خوانی نداشت به صورت دستی پیدا شده و حذف شدند. پس از تعیین موضوع‌های مرتبط با هر کشور، پارامتری به نام Topic Score معرفی می‌شود که بیان می‌کند چند درصد از کلماتی که در یک توئیت استفاده شده‌اند با کلماتی که در موضوع‌های مختلف هستند، اشتراک دارند. تعریف ریاضی آن در (۲-۱۳) آمده است که در آن، τ مجموعه‌ی توئیت‌های یک کاربر است و S مجموعه کلماتی هستند که در یک موضوع آمده‌اند و تعداد آن برای همه‌ی موضوعات ۲۵ کلمه است.

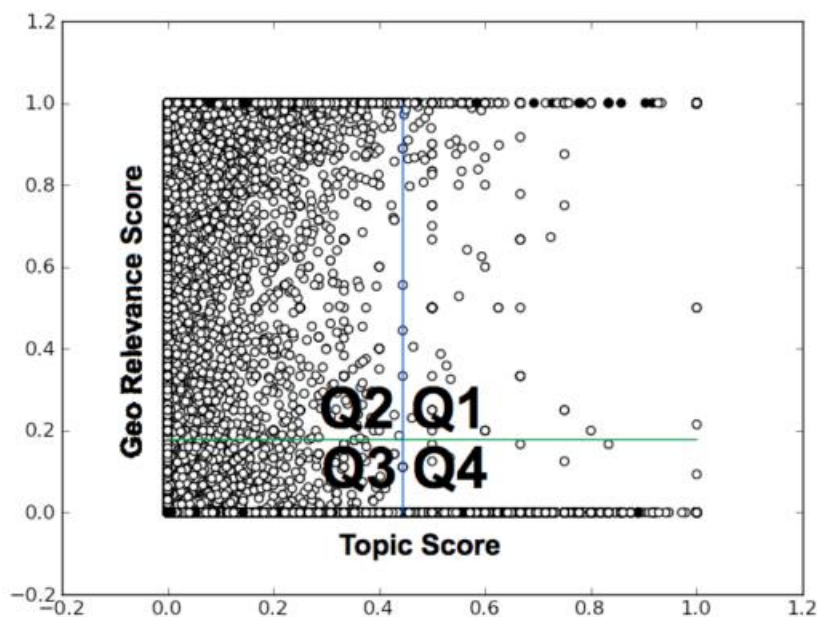
$$\text{topic_score}(S,T) = \frac{\sum_{T \in \tau} \text{sgn}(S \cap T)}{|\tau|} \quad (13-2)$$

برای تعیین دومین ویژگی توئیت‌ها، دو راه وجود دارد: (۱) استفاده از مکان‌هایی که نویسنده‌ی توئیت در مطلب خود ذکر می‌کند (در بسیاری در توئیت‌ها، کاربران ذکر می‌کنند که توئیت خود را از چه شهر/کشوری ارسال کرده‌اند)؛ (۲) استفاده از شهر/کشوری که نویسنده‌ی توئیت در پروفایل خود ذکر کرده است. برای توئیت‌هایی که در آن‌ها شهر/کشور ذکر نشده‌اند، میانگین تعداد توئیت‌هایی در نظر گرفته می‌شود که در زمان t ارسال شده‌اند و در

آن زمان، کاربر، موقعیت جغرافیایی خود را در پروفایلش ثبت کرده است. ابتدا بردار $tweet_loc \in \mathbb{R}^T$ به اندازه‌ی تعداد توئیت‌های کاربر در نظر گرفته می‌شود که در آن $tweet_loc_t = 1$ بدین معنا است که در زمان t یک توئیت ثبت شده و در همان زمان، کاربر، موقعیت جغرافیایی خود را در پروفایلش ثبت کرده است. سپس پارامتر دیگری به نام Geo Relevance Score تعریف می‌شود که میزان نزدیکی موقعیت جغرافیایی کاربر را نسبت به کشورهایی که بهار عربی در آن‌ها اتفاق افتاده است، می‌سنجد. مقدار این پارامتر در بازه‌ی $[0-1]$ است و تعریف آن در (۲-۱۴) آمده است. $\| \cdot \|_0$ بیان گر تُرم صفر است.

$$geo_rel_score(tweet_{loc}) = \frac{\|tweet_{loc}\|_0}{T} \quad (2-14)$$

حال، با توجه به دو پارامتر مشخص‌شده، کاربران بر روی یک نمودار دو بعدی، مانند شکل ۲-۴، نمایش داده می‌شوند. خطوط عمودی و افقی رسم شده در شکل ۲-۴، به ترتیب، میانگین Geo Relevance Score و Topic Score هستند.



شکل ۲-۴ نمودار رسم شده بر اساس معیارهای *Geo Relevance Score* و *Topic Score* در [۳]

چهار ناحیه‌ی مشخص‌شده در شکل ۲-۴ را می‌توان به صورت زیر تفسیر کرد:

- **Q1:** افرادی که در این گروه جای می‌گیرند کسانی هستند که هم توئیت‌های آن‌ها مرتبط با بهار عربی بوده و هم از نزدیک شاهد اتفاقاتی که در حداقل یکی از کشورهای مصر، لیبی، سوریه، تونس و یمن افتاده است، بوده‌اند. این افراد شاهدان عینی^{۷۵} می‌باشند.
- **Q2:** کاربرانی که در این چارک هستند، در متن اتفاقات بوده‌اند اما تمایلی به نوشتن در مورد تغییر و تحولات در نزدیکی خود نداشته‌اند. این کاربران، کسانی هستند که از نوشتن در مورد موضوع بهار عربی، چشم‌پوشی کرده‌اند^{۷۶}.
- **Q3:** در این بخش کاربرانی هستند که نه در متن اتفاقات مربوط به بهار عربی بوده‌اند و نه توئیت‌هایشان را به این موضوع اختصاص داده‌اند. این افراد نسبت به بهار عربی، بی‌تفاوت و بی‌احساس بوده‌اند^{۷۷}.
- **Q4:** چارک آخر اختصاص دارد به همدردهایی^{۷۸} که گرچه از لحاظ جغرافیایی به هیچ‌کدام از کشورهای مصر، لیبی، سوریه، تونس و یمن نزدیک نبوده‌اند اما توئیت‌هایشان را به بهار عربی در این کشورها اختصاص داده‌اند و از مردم آن کشورها جانب‌داری کرده‌اند.

با توجه به نتایج به دست آمده، می‌توان بیان کرد که افرادی که در چارک اول قرار می‌گیرند منبعی به مراتب قابل‌اطمینان‌تر از کاربرانی هستند در چارک‌های دیگر جای دارند؛ زیرا این افراد از نزدیک اتفاقات را رصد می‌کنند. همچنین این افراد شکل‌دهنده‌ی افکار دیگر کسانی هستند که در شبکه‌ای مانند توئیتر به دنبال مطلع شدن از اخبار یک حادثه در نقطه‌ای از دنیا هستند.

در سال ۲۰۱۰، می‌یانگ^{۷۹} و همکارانش در پژوهش [۱۳] به بررسی افراد تأثیرگذار در سه موضوع مهم که در سال ۲۰۰۹ در شبکه‌ی اجتماعی توئیتر اتفاق افتاد، پرداختند. آن‌ها در تحقیق خود به سه موضوع انتخابات ریاست جمهوری سال ۸۸ ایران، آنفولانزای خوکی و نیز مرگ خواننده معروف آمریکایی، مایکل جکسون، پرداختند. در این تحقیق از ۳

^{۷۵} Eyewitness

^{۷۶} Topic ignorant

^{۷۷} Apathetic

^{۷۸} Sympathizer

^{۷۹} Meeyoung Cha

سنجه‌ی تعداد followerها^{۸۰}، retweetها^{۸۱} و mentionها^{۸۲} برای تعیین افراد تأثیرگذار استفاده شد. این سنجه‌ها، هر یک، به طور مجزا بر روی پایگاه‌داده‌ی جمع‌آوری شده اعمال شد. نتایج زیر برای هر یک از سنجه‌ها به دست آمده است:

○ **تعداد دنبال‌کنندگان:** در این بخش، خبرگزاری‌ها، سیاست‌مداران، ورزش‌کاران و چهره‌های هنری در صدر قرار گرفتند.

○ **تعداد بازنشر توئیت‌ها:** در این سنجه، سرویس‌های محتوایی مانند Mashable^{۸۳} و همچنین سایت‌های خبری، تأثیرگذارترین کاربران در توئیت‌ر شناخته شدند.

○ **تعداد ذکر نام کاربران:** در این بخش، هنرمندان توانستند رتبه‌های برتر را از آن خود کنند.

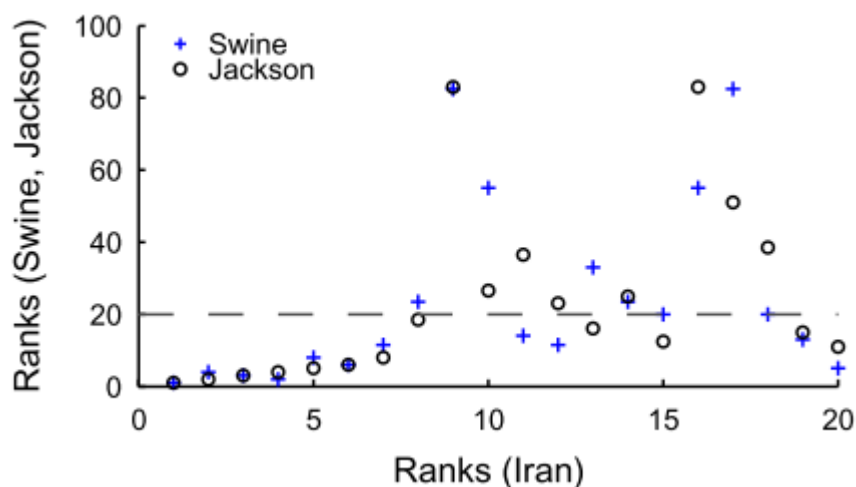
نتایج به دست آمده در هر سه سنجه نشان می‌دهد، کاربرانی که در یک موضوع تأثیرگذار شناخته شده‌اند، در دیگر موضوعات مورد بررسی نیز توانسته‌اند تأثیرگذاری خود را حفظ نمایند. شکل ۲-۵ این موضوع را در مورد تعداد retweetها نشان می‌دهد. پنج نفر اولی که در موضوع انتخابات ریاست جمهوری سال ۸۸ ایران حضور دارند، توانسته‌اند در دو موضوع دیگر نیز خود را در صدر جدول قرار دهند. از جمله‌ی این کاربران می‌توان به Mashable، CnnBrk، TweetMeme، Time و BreakingNews اشاره کرد.

^{۸۰} "دنبال‌کننده": برای توضیح در مورد follower به بخش ۳-۱ مراجعه فرمایید

^{۸۱} "بازنشر توئیت": برای توضیح در مورد retweet به بخش ۳-۱ مراجعه فرمایید

^{۸۲} "ذکر نام کاربران": برای توضیح در مورد mention به بخش ۳-۱ مراجعه فرمایید

^{۸۳} <https://mashable.com>



شکل ۲-۵ رتبه‌بندی افراد تأثیرگذار بر اساس تعداد *retweet*ها در سه موضوع مختلف [۱۳]

از جمله تحلیل‌هایی که در این تحقیق به عمل آمده است، پاسخ به این سؤال است که چگونه یک کاربر عادی توئیتر می‌تواند میزان تأثیرگذاری خود را افزایش دهد؟ پاسخ به این سؤال، طبق این پژوهش، این‌گونه بیان می‌شود که با توجه به پایگاه‌داده‌ی استفاده‌شده، کاربرانی در طول سال ۲۰۰۹ توانسته‌اند میزان تأثیرگذاری خود را افزایش دهند که تنها بر روی یک موضوع خاص تمرکز کرده و فقط در مورد آن مطلب می‌نویسند.

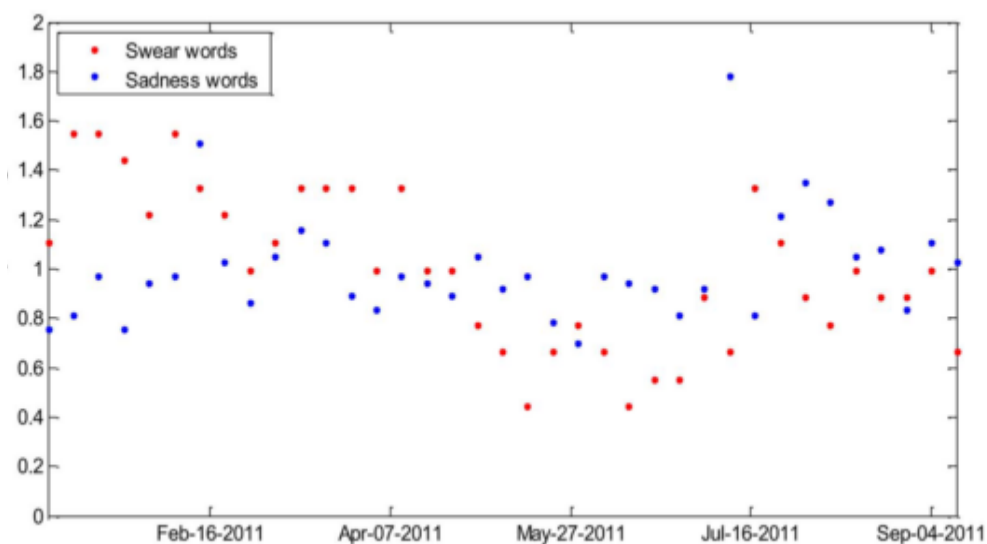
از دیگر پژوهش‌های انجام شده در زمینه‌ی تشخیص افراد تأثیرگذار در شبکه‌ی توئیتر می‌توان به [۱۵] اشاره کرد که در سال ۲۰۱۴ انجام شده است. مدل ارائه شده در این پژوهش صرفاً به تحلیل محتوایی توئیتهای می‌پردازد و ساختار شبکه را مد نظر قرار نمی‌دهد. هدف اصلی تحقیق، پیدا کردن زمان‌هایی است که نحوه‌ی نگارش توئیتهای کاربران دچار تحول می‌شود. واضح است که پس از پیدا کردن چنین زمان‌هایی، یافتن کسانی که باعث این تغییر شده‌اند نیز به عنوان افراد تأثیرگذار، کاری سهل است. لس‌سروی^{۸۴} و همکارش در این تحقیق معتقدند که برای یافتن تغییر در توئیتهای کاربران باید به دنبال تغییر در احساسات مردم در پیام‌های توئیترشان بود. به عبارت صحیح‌تر وقتی احساسات کاربران در بازه‌ای از زمان دچار تغییر می‌شود، حتماً کسی بوده است که توانسته با تأثیر خود، این تغییر را ایجاد کند. برای یافتن تغییر در احساسات کاربران توئیتر، ابتدا با استفاده از برنامه‌ی *Linguistic Inquiry and Word Count 2007*

^{۸۴} Les Servi

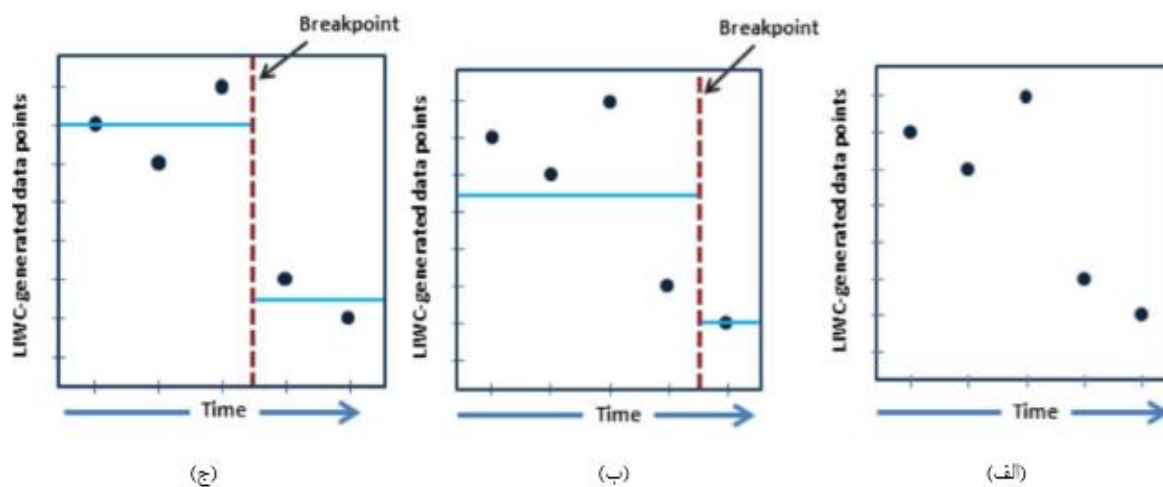
(LIWC) میزان احساسات کاربران شناسایی شده است. برنامه‌ی LIWC، خود، دارای ۸۰ کلاس مختلف است که احساسات، یکی از آنهاست. این کلاس‌ها، هر یک به چند زیرکلاس تقسیم می‌شوند. مثلاً برای کلاس احساسات، می‌توان از "ناراحتی" و "احساس مثبت" نام برد. این برنامه با دریافت متن هر توئیت، آن را با یک میزان شباهت به زیرکلاس مورد نظر، به یکی از آنها انتساب می‌دهد. شکل ۲-۶ مثالی از توئیتهای را در طول یک دوره‌ی زمانی نشان می‌دهد که توسط برنامه‌ی LIWC در زیر کلاس‌های "ناراحتی" و "عصبانیت" دسته‌بندی شده‌اند.

با استفاده از این بصری سازی داده‌ها، می‌توان مرزی را برای تغییرات احساسات در توئیتهای تعیین نمود. مشکل اصلی این کار این است که هر فرد ممکن است نظر خودش را در این مرزبندی اعمال کند؛ لذا در این تحقیق، مدلی را ارائه می‌دهد که این کار توسط ماشین صورت پذیرد. در ادامه به مثالی برای درک بهتر مدل معرفی شده خواهیم پرداخت. فرض کنید تعداد ۵ توئیت یا هر متن دیگری در اختیاردارید و توسط نرم‌افزار LIWC، احساسات هر یک مشخص شده‌اند. مدل بصری احساسات این ۵ توئیت فرضی در شکل ۲-۷ آمده است. اگر مرزی^{۸۵} که باعث تغییر در احساسات شده است، بین نقاط ۴ و ۵ باشد، میانگین داده‌ها در هر قسمت از مرز (خط آبی) به صورت شکل ۲-۷-ب، خواهد شد. اما از آنجایی که ممکن است این فرض که تغییر بین چه نقاطی است، توسط انسان اشتباه تعیین شود، اگر این مرز بین نقاط ۳ و ۴ باشد (شکل ۲-۷-ج)، خط آبی تغییر خواهد کرد.

^{۸۵} Breakpoint



شکل ۲-۶ مثالی از چند توئیت که با نرم‌افزار *LIWC* در زیرگروه‌های "ناراحتی" و "عصبانیت" دسته‌بندی شده‌اند [۱۵]



شکل ۲-۷ (الف) مثالی از تغییر احساسات در ۵ متن در طول زمان؛ (ب) ایجاد مرز جداکننده‌ی اول؛ (ج) مرز جداکننده‌ی دوم [۱۵]

برای حل این موضوع می‌توان فاصله‌ی هر نقطه تا خطوط آبی را در هر سمت از مرز جداکننده محاسبه نمود؛ هر سمتی که جمع این فاصله بیش‌تر باشد، در آن مرز، تغییر به وجود آمده است. با توجه به این تعریف، همان‌طور که در شکل ۲-۷ مشخص است، خط

جداکننده‌ی تصویر وسط بیان‌گر تغییر بیش‌تری نسبت به مرز انفصالی تصویر سمت راست است.

در یکی دیگر از تحقیقاتی که در سال ۲۰۱۴ صورت گرفت، ایرما مترا^{۸۶} از دانشگاه پلی‌تکنیک میلان^{۸۷} ایتالیا در پایان‌نامه‌ی ارشد خود [۱۶] به تشخیص افراد تأثیرگذار در شبکه‌ی توئیتر پرداخت. وی در پایان‌نامه‌ی خود ابتدا با استفاده از رابط کاربردی برنامه-نویسی توئیتر^{۸۸} داده‌های مورد نظر خود را استخراج کرده است. برای استخراج داده‌ها، کاربران به صوت تصادفی انتخاب‌شده و سپس کاربرانی در پایگاه‌داده مورد مطالعه قرار گرفته‌اند که در توئیتر فعال بوده‌اند. تشخیص فعال بودن یک کاربر با دو ویژگی زیر مشخص می‌شود:

○ بسامد توئیتهای^{۸۹}: از آنجایی که بسیاری از کاربران توئیتر را هرزنامه-نویسان تشکیل می‌دهند، در پایان‌نامه‌ی مذکور، کاربرانی انتخاب می-شوند که رابطه‌ی (۲-۱۵) برای آن‌ها صادق باشد. در این رابطه، $Date_{registered}$ تاریخی است که کاربر در توئیتر حساب کاربری باز کرده است و $Date_{data_collection}$ تاریخی است که پایگاه‌داده‌ی مورد نظر در آن زمان جمع‌آوری شده است.

$$(0.000000002) < \frac{\sum_{Date_{registered}}^{Date_{data_collection}} Days}{\#tweets_total} < (2) \quad (15-2)$$

○ تازگی آخرین توئیت^{۹۰}: این فاکتور که توسط نویسنده‌ی پایان‌نامه تعریف شده است بیان می‌کند که آخرین توئیت کاربر باید حتماً در بازه‌ی بین تاریخ جمع‌آوری پایگاه‌داده و ۱۲۰ روز قبل از آن باشد. پس از استخراج کاربران فعال، فاکتورهای زیر برای هر کاربر به منظور تعیین افراد تأثیرگذار انتخاب شدند:

○ تعداد توئیتهای

^{۸۶} Irma Metra

^{۸۷} Polytechnic University of Milan

^{۸۸} در بخش ۳-۱ معرفی خواهد شد

^{۸۹} Tweet frequency

^{۹۰} Recency of last tweet

- تعداد توئیت‌هایی از کاربر مورد نظر که مورد پسند واقع شده
- تعداد لیست‌هایی که نام کاربر در آن‌ها آمده یا آن‌ها را ساخته
- تعداد دنبال کنندگان
- تعداد دنبال شوندگان توسط کاربر مورد نظر

سپس ۱۰۰ توئیت آخر هر کاربر برای انجام پژوهش استخراج شده و فاکتورهای زیر نیز برای هر یک از توئیت‌ها در نظر گرفته شد:

- تعداد دفعات بازنشر توئیت
- تعداد URL‌های موجود در توئیت
- تعداد هشتگ‌های توئیت
- تعداد دفعاتی که توئیت مورد نظر مورد پسند واقع شده
- تعداد mention‌های موجود در توئیت

در نهایت، پس از انتخاب فاکتورهای مورد نیاز برای انجام پژوهش، با استفاده از الگوریتم AHP [۱۷]، به هر یک از فاکتورها، وزنی اختصاص داده شده است. در مرحله‌ی آخر همه‌ی فاکتورها در وزن‌های خود ضرب شده و با یکدیگر جمع می‌شوند. عدد به دست آمده، بیان‌گر میزان تأثیرگذاری کاربر مورد نظر در شبکه‌ی توئیتر است.

۲-۳- مدل‌های عمومی

یوژیان^{۹۱} در سال ۲۰۱۴ با استفاده از روش TOPSIS^{۹۲} توانست مدلی برای تشخیص افراد تأثیرگذار در یک شبکه‌ی اجتماعی در [۱۸] ارائه دهد. TOPSIS روشی است در فرآیند تصمیم‌گیری چند متغیره^{۹۳} مورد استفاده قرار می‌گیرد. در پژوهش صورت گرفته، مدل ارائه شده در ۴ شبکه‌ی گرافی به کار گرفته شده

^{۹۱} Yuxian Du

^{۹۲} Technique for Order Performance by Similarity to Ideal Solution

^{۹۳} Multiple attribute decision making (MADM)

است: الف) شبکه‌ی پروازهای ایالات متحده (هر پرواز بیان گر یک یال^{۹۴} است و مبدأ و مقصد پرواز، گره‌های^{۹۵} گراف می‌باشند؛ ب) شبکه‌ی خطوط هوایی^{۹۶}؛ ج) شبکه‌ی رایانامه‌های^{۹۷} ارسال شده بین اعضای University Rovira i Virgili (Tarragona) [۱۹]؛ د) شبکه‌ی مسابقات فوتبال بین باشگاهی [۲۰]. در این مقاله با استفاده از ۴ سنجی درجه، نزدیکی، بینابینی و بردار ویژه، افراد تأثیرگذار مشخص شده‌اند. در دو پایگاه داده‌ی اول که اطلاعات از طریق یک مسیر^{۹۸} جریان پیدا می‌کند، از سنجی‌های درجه، نزدیکی و بینابینی استفاده شده است و در دو پایگاه داده‌ی آخر که اطلاعات در شبکه‌ی آن‌ها از طریق یک گشت^{۹۹} رد و بدل می‌شود، سنجی‌های درجه، نزدیکی و بردار ویژه مورد استفاده قرار گرفته‌اند.

یکی از مهم‌ترین مسائلی که در مواجهه با سنجی‌های نزدیکی و مرکزیت پیش می‌آید، پیچیدگی محاسباتی آن‌ها است. این موضوع باعث می‌شود نتوان از آن‌ها در شبکه‌های بسیار بزرگ استفاده نمود. سنجی درجه نیز نمی‌تواند معیار خیلی خوبی برای بیان تأثیرگذار بودن یا نبودن افراد در یک شبکه باشد. این عوامل باعث شده‌اند که دوآنبینگ و همکارانش در سال ۲۰۱۲ پژوهشی را انجام دهند و معیار جدیدی به نام مرکزیت محلی^{۱۰۰} را در [۲۱] معرفی نمایند. سنجی فوق از روابط (۲-۱۶) و (۲-۱۷) قابل محاسبه است که در آن، $\Gamma(u)$ مجموعه‌ای از نزدیک‌ترین همسایه‌های (همسایه‌های سطح ۱) گره‌ی u هستند و $N(w)$ تعداد همسایه‌های سطح ۱ و ۲ گره‌ی w می‌باشند.

$$Q(u) = \sum_{w \in \Gamma(u)} N(w) \quad (۱۶-۲)$$

$$C_L(v) = \sum_{u \in \Gamma(v)} Q(u) \quad (۱۷-۲)$$

مدل ارائه شده را با ذکر یک مثال توضیح خواهیم داد. فرض کنید شکل ۲-۸ شبکه‌ی مورد نظر ما باشد. در این شکل، گره‌ی ۱، هشت همسایه‌ی سطح ۱ (گره‌های ۲ تا ۹) و یک همسایه‌ی سطح ۲ که همان گره شماره ۱۰ باشد، دارد؛ در نتیجه داریم $N(1) = 9$. با توجه به رابطه‌ی (۲-۱۶)، $Q(1) = N(2) + N(3) + N(4) + N(5) + N(6) + N(7) + N(8) + N(9) = 67$ به همین روش، می‌توانیم Q ها را برای هر گره محاسبه کنیم.

^{۹۴} Edge

^{۹۵} Node

^{۹۶} <http://wiki.gephi.org/index.php/Datasets>

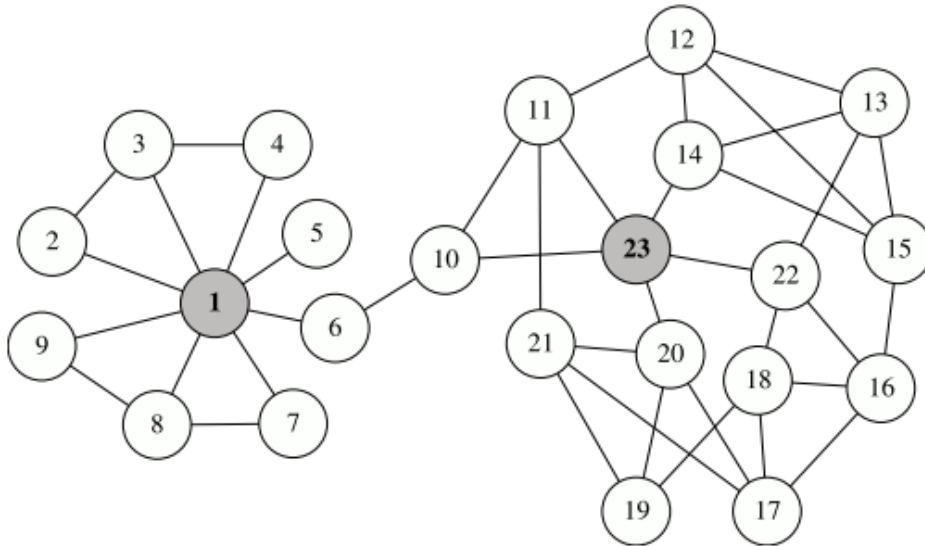
^{۹۷} Email

^{۹۸} Path

^{۹۹} Walk

^{۱۰۰} Local centrality

برای ارزیابی مدل ارائه شده، ۴ پایگاه داده در نظر گرفته شده‌اند: الف) وبلاگ‌ها [۲۲]؛ ب) دانشمندان [۲۳]؛ ج) روترها^{۱۱} [۱۹]؛ د) رایانامه‌ها [۲۰]. مقایسه‌ی سنجی مرکزیت محلی با دیگر سنجه‌ها در جدول ۱-۲ آمده است.



شکل ۲-۸ مثالی از یک شبکه برای تشریح مدل مرکزیت محلی [۲۱]

جدول ۲-۷- مقایسه‌ی رتبه‌بندی گره‌ها با استفاده از سنجه‌های محلی (L)، درجه (D)، نزدیکی (C) و بینابینی (B) در شبکه‌های (از چپ به راست) وبلاگ‌ها، روترها، دانشمندان و ایمیل‌ها [۲۱]

B	C	D	L
۲	۳	۱	۱
۱۰	۴	۳	۲
۱	۱	۲	۳
۲۲	۴۰	۴	۴
۳	۲	۵	۵
۶۱	۱۱	۱۹	۶
۱۵	۱۹	۷	۷
۸	۵	۶	۸
۱۶	۲۱	۹	۹
۳۶	۳۳	۱۲	۱۰

B	C	D	L
۱۲	۱۹	۱	۱
۶	۷	۲	۲
۵۰	۷۷	۴	۳
۲۱	۸۱	۸	۴
۱۰۱	۸۵	۲۹	۵
۱۳۵	۸۶	۴۴	۶
۱۳۶	۸۷	۴۵	۷
۱۳۷	۸۸	۴۶	۸
۱۳۸	۸۹	۴۷	۹
۱۵	۲۲	۳۰	۱۰

B	C	D	L
۳۱	۱۶	۵	۱
۲	۲	۹	۲
۴	۷	۷	۳
۳۶	۱۹	۶	۴
۱	۱	۱۱	۵
۵	۴	۱۵	۶
۲۴	۹	۲۰	۷
۵۲	۲۰	۱۶	۸
۲۰	۸	۳۷	۹
۴۷	۲۱	۳۱	۱۰

B	C	D	L
۳	۳	۱	۱
۴۵	۱۱	۱۵	۲
۱۶	۷	۳	۳
۶۶	۱۵	۳۷	۴
۵۰	۱۶	۱۶	۵
۷۴۱	۵۷	۸۷	۶
۸۳۲	۶۳	۹۶	۷
۱۴۰	۲۲	۸۸	۸
۱۸۴	۵۶	۱۰۹	۹
۷۰۷	۷۸	۱۳۵	۱۰

^{۱۱} Routers

^{۱۲} Carson Kai-Sang Leung

downward closure است. در این مدل متغیری به نام Prominence تعریف می‌شود که در واقع میزان محبوبیت یک فرد را مشخص می‌کند. این مقدار می‌تواند تعداد دوستان یک فرد، میزان فعالیت وی در شبکه و ... باشد. برای محاسبه‌ی میزان Prominence برای یک گروه، میانگین Prominence تمامی افراد گروه محاسبه می‌شود. متغیر دیگری که در این مدل مورد استفاده قرار می‌گیرد، تعداد دفعاتی است که یک فرد در گروه‌های مختلف ظاهر شده است (Frequency). ضرب این دو متغیر تعریف شده، بیان می‌کند آیا یک فرد می‌تواند در یک گروه به عنوان کاربر تأثیرگذار وارد شود یا خیر. نتایج به دست آمده از اعمال این مدل بر روی ۴ پایگاه داده‌ی T10I4D100K^{۱۰۳}، mushroom^{۱۰۴}، pumsb^{۱۰۵} و kosarak^{۱۰۶} حاکی از تولید درختی کم حجم است. از دیگر نتایج به دست آمده می‌توان به سرعت این مدل در پردازش اطلاعات و نیز مقیاس‌پذیری^{۱۰۷} آن در رابطه با ماتریس‌های متراکم^{۱۰۸} و پراکنده^{۱۰۹} اشاره کرد.

۲-۴- بررسی ابزارهای رتبه‌بندی و تحلیلی

نفوذ شبکه‌های اجتماعی به زندگی خصوصی افراد و تبلیغاتی که شرکت‌های مختلف برای محصولات خود در این شبکه‌ها انجام می‌دهند و نیز تأثیرپذیری مردم در مسائل سیاسی، اجتماعی، فرهنگی و ... از ستاره‌های سینما و تلویزیون، سیاستمداران و ورزشکاران مطرح دنیا، مسائلی هستند که باعث شده‌اند سایت‌هایی برای رتبه‌بندی کاربران شبکه‌های اجتماعی راه‌اندازی شوند. این سایت‌ها با استفاده از یک سری الگوریتم‌های خاص، که غالباً منتشر نمی‌شوند [۲۶]، کاربران را رتبه‌بندی می‌کنند. این رتبه‌بندی با توجه به تفاوت در معیارهای تأثیرگذار بودن کاربران و نیز الگوریتم‌های مختلف، برای هر سایت ممکن است متفاوت باشد. مثلاً در مورد حساب کاربری آقای دکتر ظریف، وزیر امور خارجه کشورمان، در دو سایت Klout^{۱۱۰}

^{۱۰۳} <http://www.almaden.ibm.com/cs/quest>

^{۱۰۴} <http://fimi.ua.ac.be/data/mushroom.dat>

^{۱۰۵} <http://fimi.ua.ac.be/data/pumsb.dat>

^{۱۰۶} <http://fimi.ua.ac.be/data/kosarak.dat>

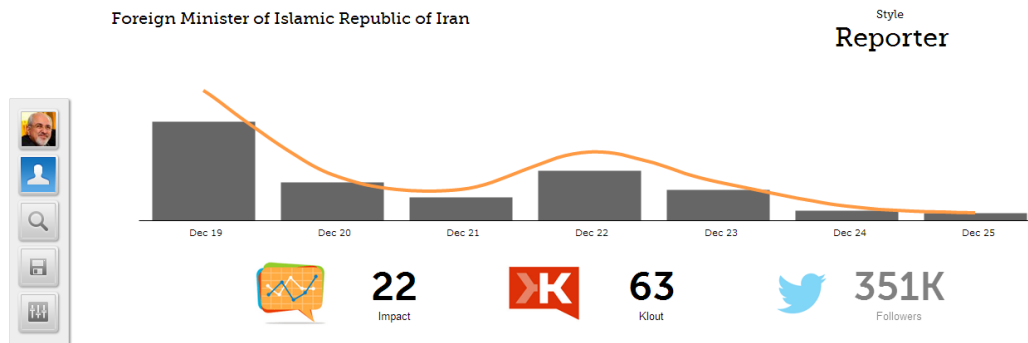
^{۱۰۷} Scalability

^{۱۰۸} Dense matrix

^{۱۰۹} Sparse matrix

^{۱۱۰} <https://klout.com>

و Twitalyzer^{۱۱۱} که هر دو عددی بین ۰ تا ۱۰۰ را به عنوان میزان تأثیرگذاری در توئیتر اختصاص می‌دهند (۰=کم‌ترین امتیاز؛ ۱۰۰=بیش‌ترین امتیاز)، به صورت شکل ۹-۲ هست.



شکل ۹-۲ عدد تأثیرگذاری آقای دکتر ظریف، وزیر امور خارجه کشورمان، در دو سایت Klout (عدد ۶۳) و Twitalyzer (عدد ۲۲)

در این قسمت، به معرفی برخی از ابزارهای رتبه‌بندی کاربران شبکه‌های اجتماعی و مخصوصاً توئیتر اشاره خواهیم کرد:

Klout - ۱-۴-۲

از معروف‌ترین سایت‌های رتبه‌بندی می‌توان به Klout اشاره کرد. این سایت با بیش از ۶۲۰ میلیون کاربر رتبه‌بندی شده [۴۷]، توانسته است بازار کار را از دیگر رقبا برپاید. از نکات مثبت این ابزار می‌توان به پشتیبانی آن از چندین شبکه‌ی اجتماعی نام برد. این سایت می‌تواند کاربران شبکه‌های فیس‌بوک، توئیتر، لینکدین، اینستاگرام، گوگل پلاس، فوراسکوئر^{۱۱۲} و یوتیوب را رتبه‌بندی کند. از دیگر نکات Klout می‌توان به امکان بررسی میزان تأثیرگذاری در سه ماهه‌ی اخیر اشاره کرد، به طوری که کاربران با نگاه کردن به نمودار رسم شده، می‌توانند روند صعودی، نزولی یا ثابت خود را در هر یک از شبکه‌های یاد شده دنبال کنند. شکل ۱۰-۲ روند رو به رشد یکی از کاربران اینستاگرام با نام کاربری behzad.rezaie11 را در Klout نشان می‌دهد. از دیگر ویژگی‌های این سایت می‌توان به پیشنهادهایی اشاره کرد که می‌تواند باعث افزایش تأثیرگذاری کاربر

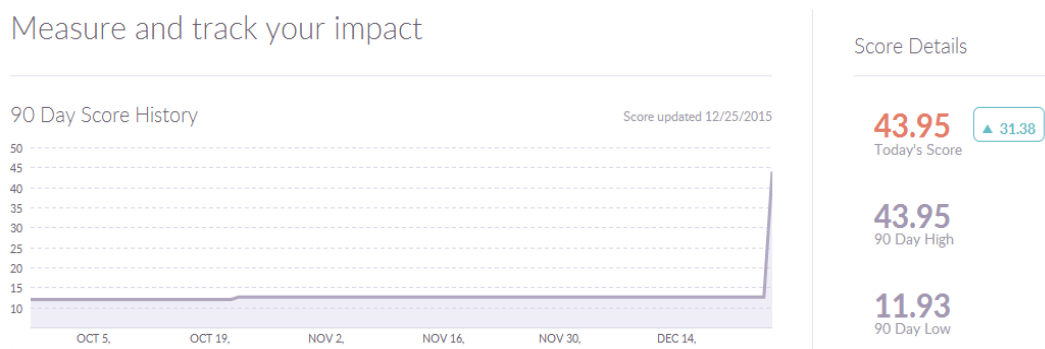
^{۱۱۱} <http://twitalyzer.com>

^{۱۱۲} <https://foursquare.com>

شود. از جمله پیشنهادهای این ابزار می‌توان به کسانی اشاره کرد که دنبال کردن آن‌ها باعث افزایش تأثیرگذاری می‌شود. یکی دیگر از ویژگی‌های Klout این است که کاربران می‌توانند با استفاده از آن، مطالب خود را در شبکه‌های اجتماعی پشتیبانی شده توسط این سایت، پست کنند.

سنجه‌های مورد استفاده در Klout را می‌توان به ۳ بخش تقسیم کرد:

- **True reach**: این عامل بیان‌گر تعداد مخاطبانی است که با لایک کردن یا بازنشر توئیتهای کاربر، با وی در ارتباط هستند. [۲۶]
- **Amplification probability**: این عامل مشخص‌کننده‌ی تعداد لایک‌ها، نظرات و بازنشرهای یک توئییت است. [۲۶]
- **Network score**: این عامل بیان‌گر میزان تأثیرگذاری دوستان یک کاربر یا کسانی است که پست‌های وی را مورد پسند قرار می‌دهند یا بازنشر می‌کنند. [۲۶]



شکل ۲-۱۰ مثالی از روند رو به رشد تأثیرگذاری کاربری با نام کاربری *behzad.rezaie11* در شبکه‌ی اینستاگرام. نمودار توسط ابزار *Klout* رسم شده است

از نکات بارز Klout برای توسعه‌دهندگان نرم‌افزار می‌توان به رابط کاربردی برنامه‌نویسی آن اشاره کرد^{۱۱۳}. با استفاده از رابط کاربردی برنامه‌نویسی فراهم شده، می‌توان عدد تأثیرگذاری کاربران توئیتر را با ارسال نام کاربری‌شان به دست آورد.

Twitalyzer -۲-۴-۲

^{۱۱۳} <https://klout.com/s/developers/home>

از دیگر سایت‌های رتبه‌بندی، می‌توان به Twitalyzer اشاره کرد. استفاده از ابزارهای تحلیلی این سایت تا اول اکتبر ۲۰۱۲ به صورت رایگان بوده است، اما اکنون برای استفاده از این ابزارها، باید هزینه پرداخت شود. این سایت امکانی را در اختیار کاربران خود قرار می‌دهد که با جست‌وجوی نام کاربری کاربران توئیتر، بتوان میزان تأثیرگذاری آن‌ها را به صورت کمی مشاهده کرد، این عدد را با عدد به دست آمده از Klout مقایسه کرد و نیز روند آن را در یک بازه‌ی ۹۰ روزه مورد مطالعه قرار داد. شکل ۲-۱۰ مثالی از نمودارهای رسم شده توسط این ابزار است. این ابزار تنها کاربران توئیتر را مورد سنجش قرار می‌دهد.

۲-۴-۳- Twittercounter^{۱۱۴}

یکی دیگر از ابزارهای موجود در رابطه با رتبه‌بندی کاربران توئیتر، Twittercounter است. این سایت کاربران توئیتر را از لحاظ تعداد دنبال‌کنندگان، دنبال‌شوندگان و توئیتهای نیز رتبه‌بندی می‌کند. از دیگر نکات جالب این ابزار می‌توان به امکان فیلتر کردن نتایج جست‌وجو بر اساس منطقه‌ی جغرافیایی یا زمینه‌ی کاری کاربران نام برد؛ مثلاً می‌توان ۱۰۰ کاربر برتر ایرانی که بیش‌ترین تعداد دنبال‌کننده را دارند^{۱۱۵} و یا ۱۰۰ ورزش‌کار برتر دنیا از لحاظ تعداد دنبال‌کننده^{۱۱۶} را مشاهده نمود.

برخی از سرویس‌های ارائه شده در Twittercounter رایگان و برخی دیگر با پرداخت هزینه قابل دسترس هستند. پس از وصل شدن به حساب کاربری توئیتر خود از طریق امکانی که در سایت وجود دارد، این ابزار تعداد دنبال‌کنندگان، دنبال‌شوندگان و توئیتهای کاربر را در طول زمان بر روی نموداری به وی نشان می‌دهد. همچنین، رتبه‌ی کلی کاربر را از لحاظ تأثیرگذاری می‌توان در صفحه‌ی خانگی وی مشاهده نمود. شکل ۲-۱۱ نمونه‌ای از رتبه‌ی کلی، تعداد دنبال‌کنندگان و دنبال‌شوندگان را در طول زمان برای فردی با نام کاربری @Behzad28 در توئیتر نشان می‌دهد.

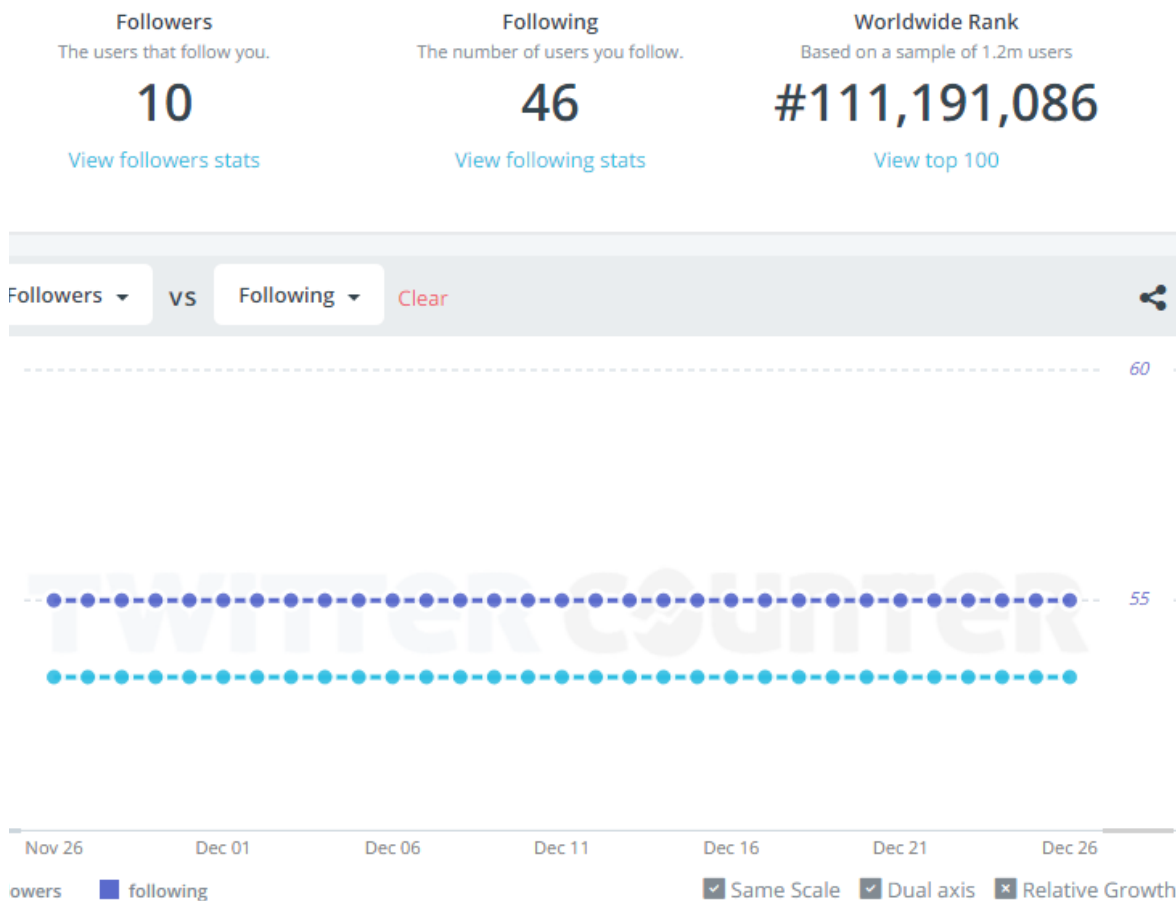
^{۱۱۴} <http://twittercounter.com>

^{۱۱۵} <http://twittercounter.com/pages/100/iran>

^{۱۱۶} <http://twittercounter.com/pages/100/sports-athlete>

از دیگر امکانات این ابزار، که به صورت پولی می‌باشند، می‌توان به موارد زیر اشاره کرد:

- امکان گرفتن گزارش از نمودارهای ارائه شده به صورت فایل‌های pdf و CSV
- امکان مشاهده‌ی بازنشر توئیت‌ها، mentionها و تعداد دفعات مورد پسند واقع شدن یک توئیت
- پیشنهاد در مورد بهترین زمان برای ارسال توئیت
- مقایسه‌ی میزان تأثیرگذاری خود با کاربران دیگر در توئیت
- امکان مشاهده‌ی کسانی که خود را از لیست دنبال‌کنندگان کاربر خارج کرده‌اند
- امکان دستیابی به اطلاعاتی مانند تعداد دنبال‌کنندگان، دنبال‌شوندگان، توئیت‌ها و ... برای هر یک از کاربران توئیت



شکل ۲-۱۱ نمونه‌ای از آمار کاربر @Behzad28 ارائه شده در *Twittercounter*

این ابزار تنها برای کاربران توئیتر سرویس دهی می‌کند.

۲-۴-۴-۱۱۷ Hootsuite

Hootsuite از جمله ابزارهایی است که بیش‌تر در حوزه‌ی تحلیل داده‌های شبکه‌های اجتماعی فعالیت دارد. این ابزار از شبکه‌های توئیتر، فیس‌بوک، گوگل پلاس، لینکدین، اینستاگرام و وردپرس^{۱۱۸} پشتیبانی می‌کند. کاربران می‌توانند هم‌زمان به چندین

^{۱۱۷} <https://hootsuite.com>

^{۱۱۸} <https://wordpress.com>

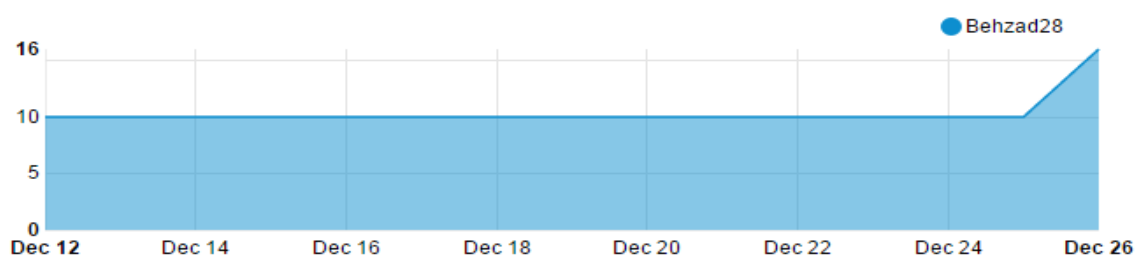
حساب کاربری خود در شبکه‌های مختلف متصل شده و در یک لحظه تمامی آن‌ها را مدیریت کنند. شکل ۲-۱۲ نمونه‌ای از مدیریت حساب‌های کاربری در شبکه‌های لینکدین و توئیتر را نشان می‌دهد. در بخش دیگری از سایت، تحلیل داده‌ها صورت می‌پذیرد. با توجه به این‌که برای دسترسی به بسیاری از سرویس‌های این ابزار باید هزینه‌ی آن‌ها پرداخت شود، تنها داده‌ای که به صورت رایگان در اختیار قرار می‌گیرد، آماری از تعداد افراد دنبال‌کننده، دنبال‌شونده، لیست‌ها و تغییرات تعداد دنبال‌کنندگان در طول زمان است. شکل ۲-۱۳ این آمار را برای کاربر Behzad28 نشان می‌دهد. از دیگر تحلیل‌های این سایت می‌توان به تعداد بازنشر توئیت‌ها، mentionها و لایک‌ها به همراه جزئیات آن‌ها نام برد که برای استفاده از این گزارش‌ها باید هزینه‌های مربوطه پرداخت گردد. از دیگر امکانات Hootsuite می‌توان به ارائه‌ی لیست دنبال‌کنندگان نام برد. در این لیست آمار مختصری در مورد این کاربران به همراه امتیاز تأثیرگذاری آن‌ها در Klout نیز آمده است.

The screenshot displays the Hootsuite interface for user Behzad28. The main feed shows four tweets from Liverpool FC (LFC) dated December 26, 2015. The tweets discuss Christian Bolasie's performance and the team's strategy. The sidebar on the right includes navigation buttons for Home, Mentions, Retweets, Followers, Inbox, Lists, Likes, Outbox, My Tweets, Search, and Scheduled. The top right corner has an 'Add a stream' button and a dropdown menu for the user profile.

شکل ۲-۱۲ نمونه‌ای از مدیریت چند حساب کاربری در شبکه‌های گوناگون توسط ابزار Hootsuite

Followers	16
Following	55
Listed	0

lower Growth



POWERED BY **TWITTER COUNTER**

شکل ۱۳-۲ نمونه‌ای از آمار ارائه شده توسط *Hootsuite* برای کاربر *@Behzad28*

همان‌طور که در ابتدای این بخش هم ذکر شد، از جمله مهم‌ترین مسائلی که در مورد ابزارهای رتبه‌بندی وجود دارد، در دسترس نبودن الگوریتم‌های مورد استفاده‌ی آن‌ها است. لذا کاربری که مثلاً در Klout، امتیاز ۵۴ می‌گیرد، علت این امتیاز را نمی‌داند و یا حتی نمی‌داند چگونه می‌تواند امتیاز خود را در این ابزار افزایش دهد؛ لذا در تحقیقی که در [۲۶] در دانشگاه مالاگای^{۱۱۹} اسپانیا توسط دل کامپواویلا^{۱۲۰} و همکارانش در سال ۲۰۱۳ انجام شد، ابزارهای Klout، PeerIndex^{۱۲۱} و TwitterGrader^{۱۲۲} مورد مطالعه قرار گرفتند. در این تحقیق، برای بررسی عوامل تأثیرگذاری در هر یک از ابزارهای یاد شده، از درخت تصمیم (C4.5) و الگوریتم رگرسیون (REPTREE و M5P) در نرم‌افزار Weka استفاده شده است. شکل ۲-۱۴ درخت‌های رسم

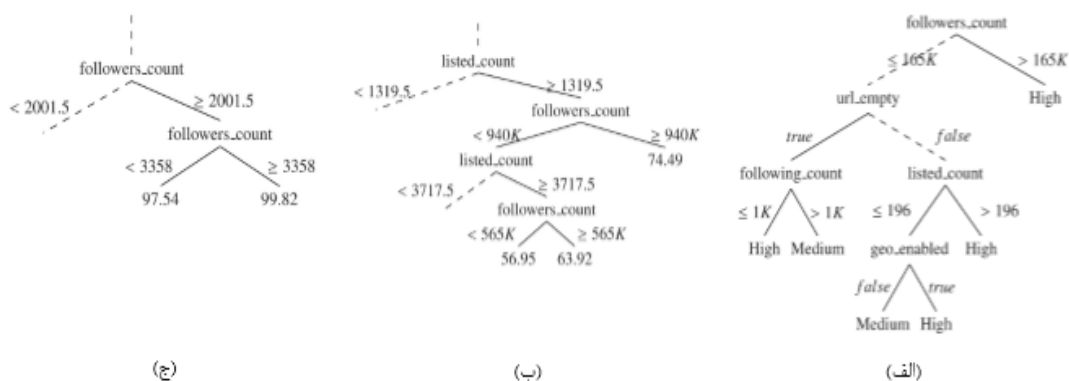
^{۱۱۹} University of Malaga

^{۱۲۰} Del Campo-Avila

^{۱۲۱} <https://brandwatch.com/peerindex-and-brandwatch>

^{۱۲۲} <http://twitter.grader.com>

شده برای هر سه ابزار را نمایش می‌دهد. همان‌طور که در شکل مشخص شده است، تعداد دنبال‌کنندگان نقش بسیار پررنگی را در همه‌ی سایت‌های یاد شده بازی می‌کند. برخی از ویژگی‌های پروفایل مانند ذکر مکان جغرافیایی نیز از اهمیت ویژه‌ای در Klout برخوردار است.



شکل ۲-۱۴ ویژگی‌های مهم استخراج شده از ابزارهای *Klout*، *PeerIndex* و *TwitterGrader*. الف) درخت تصمیم C4.5 رسم شده برای ابزار *Klout*؛ ب) درخت رگرسیون *REPTREE* رسم شده برای ابزار *PeerIndex*؛ ج) درخت رگرسیون *REPTREE* رسم شده برای ابزار *TwitterGrader* [۲۶]

از جمله نتایج این پژوهش می‌توان به اهمیت ساختار شبکه‌ای کاربر برای بهبود میزان تأثیرگذاری در ابزار *TwitterGrader* اشاره کرد، حال آن‌که *Klout* و *PeerIndex* از سنجه‌های دیگری مانند تعداد لیست‌هایی که نام کاربر در آن‌ها قرار دارد و یا ذکر مکان جغرافیایی و ... استفاده می‌کنند.

۳- روش پیشنهادی تعیین افراد تاثیرگذار در شبکه اجتماعی

در این بخش از پایان‌نامه به معرفی روش پیشنهادی خواهیم پرداخت. ابتدا شبکه‌ی اجتماعی توئیتر به همراه برخی از ویژگی‌ها و اصطلاحات آن معرفی خواهد شد. سپس ابزارها و پایگاه‌داده‌ی استفاده‌شده در تحقیق مورد بررسی قرار خواهند گرفت. در بخش آخر این فصل هم روش ارائه شده در تحقیق خواهد آمد.

۳-۱- معرفی توئیتر

توئیتر، به عنوان یکی از محبوب‌ترین شبکه‌های اجتماعی، در این تحقیق مورد مطالعه قرار گرفته است. در این بخش، به معرفی برخی از مهم‌ترین اصطلاحات رایج در آن خواهیم پرداخت.

۳-۱-۱- توئیت

یکی از مهم‌ترین ویژگی‌های این پرنده‌ی خوش‌صدای معروف، محدودیت این رسانه در تولید پیام‌ها است. پیام‌های توئیتر که "توئیت"^{۱۳۳} نام دارند، می‌توانند دارای حداکثر ۱۴۰ کاراکتر باشند [۲۷]. این ویژگی منحصربه‌فرد باعث شده است کاربران این شبکه به صورت موجز و مختصر پیام مد نظر خود را به دیگر مخاطبان توئیتر برسانند. این در حالی است که در بسیاری از شبکه‌های اجتماعی دیگر نظیر فیس‌بوک، لینکدین، لاین و ... کاربران، محدودیتی در رابطه با تعداد کاراکترهایی که می‌توانند در پیام خود ارسال کنند، ندارند.

۳-۱-۲- دنبال‌کننده^{۱۳۴}

از دیگر ویژگی‌های منحصربه‌فرد این شبکه‌ی اجتماعی می‌توان به امکان دنبال کردن^{۱۳۵} دیگران نام برد. این ویژگی بر خلاف شبکه‌های دیگر مانند فیس‌بوک، یک ارتباط دوطرفه نیست [۲۷]. در این نوع ارتباط اگر کاربر A، کاربر B را دنبال کند، اصطلاحاً یکی از دنبال‌کننده‌ی B است. عمل دنبال کردن A، بدون اجازه‌ی B امکان‌پذیر است. این ارتباط از آن‌جایی یک‌طرفه است که در این سناریو، A دنبال-کننده‌ی B است؛ اگر کاربر B هم A را دنبال کند، این رابطه دوطرفه می‌شود. این در حالی است که در شبکه‌های دیگر مانند فیس‌بوک و لینکدین، ابتدا کاربر A برای ارتباط داشتن با B، درخواست دوستی می‌دهد و اگر B این درخواست را قبول کند، B نیز یکی از دوستان A می‌شود. اگر ارتباط کاربران در توئیتر و فیس‌بوک را بخواهیم

^{۱۳۳} Tweet

^{۱۳۴} Follower

^{۱۳۵} Follow

به صورت یک گراف رسم کنیم، گراف دوستی در شبکه‌ی توئیتر، یک گراف جهت‌دار و گراف دوستی در فیس‌بوک و لینکدین، یک گراف بدون جهت خواهد بود.

توئیتر از جمله اولین شبکه‌های اجتماعی است که مفهوم هشتگ را به کاربران این شبکه‌ها معرفی کرد. این مفهوم باعث می‌شود کاربران بتوانند موضوع (موضوعات) توئیت خود را مشخص نمایند. برای استفاده از این قابلیت، کافی است قبل از عنوان موضوع مورد نظر، علامت # گذاشته شود. کاربران غالباً در انتهای توئیت خود، با استفاده از هشتگ، موضوع آن را مشخص می‌نمایند. مثلاً اگر توئیت ارسالی در مورد دانشگاه باشد، کافی است #University در پیام مورد نظر درج شود. اگر موضوع توئیت بیش از یک کلمه باشد، باید به نحوی آن کلمات را به یکدیگر متصل کرد؛ زیرا تنها یک کلمه بعد از # می‌تواند نوشته شود. به عنوان مثال برای درج نام دانشگاه صنعتی شاهرود به عنوان یکی از موضوعات یک توئیت، باید این موضوع به صورت `#Shahrood_University_of_Technology` یا `#ShahroodUniversityOfTechnology` ثبت گردد. امروزه هشتگ‌ها عضو جداناپذیر شبکه‌های اجتماعی شده‌اند. نه تنها توئیتر، بلکه شبکه‌های دیگر نیز مانند اینستاگرام و فیس‌بوک از این قابلیت بهره برده‌اند. استفاده‌ی بسیار زیاد کاربران شبکه‌های اجتماعی از هشتگ‌ها باعث شده است سایت‌هایی مانند `hashtagify`^{۱۲۷} به بررسی و تحلیل هشتگ‌ها بپردازند. از جمله اتفاقاتی که در سال‌های اخیر در مورد هشتگ‌ها افتاده است، خلق هشتگ‌های جدید برای اتفاقات و حوادث مهم دنیا است. مثلاً در مورد حملات تروریستی سال ۲۰۱۵ در پاریس، هشتگ‌هایی مانند `#PrayForParis` و `#ParisAttacks` به طور بسیار سریع ایجاد شدند و به شکل وسیع در توئیتر مورد استفاده قرار گرفتند. استفاده از هشتگ‌ها باعث می‌شود جست‌وجو در توئیت‌ها بسیار سریع‌تر و روان‌تر صورت پذیرد. در قسمت معرفی رابط کاربردی برنامه‌نویسی توئیتر به این موضوع اشاره خواهد شد.

^{۱۲۶} Hashtag

^{۱۲۷} <http://hashtagify.me>

۳-۱-۴- ذکر نام کاربران^{۱۲۸}

فرض کنید کاربری مانند A توئیتی را منتشر کرده است. کاربران بسیاری مانند B، C و ... در زیر آن نظر گذاشته‌اند. اگر فرد A بخواهد جواب هر یک از این نظرات را بدهد، مشخص نمی‌شود کدام نظر، برای کدام کاربر گذاشته شده است. در این قسمت، توئیتر ویژگی‌ای به نام mention را معرفی کرده است که کاربران می‌توانند نام کاربر دیگری را در توئیت یا نظر خود بیاورند و وی را مخاطب خود قرار دهند. برای استفاده از این امکان، کافی است ابتدا کاراکتر @ و بلافاصله پس از آن، نام کاربری فرد مورد نظر قرار بگیرد. مثلاً برای نام بردن از فردی با نام کاربری abc، کافی است عبارت @abc در توئیت قرار بگیرد. پس از ارسال توئیت، کاربر abc پیامی مبنی بر ذکر نام کاربری وی در یکی از توئیتهای ارسالی در توئیتر دریافت می‌کند. این ویژگی در شبکه‌های دیگر مانند فیس‌بوک و اینستاگرام tag نام دارد.

۳-۱-۵- بازنشر توئیت^{۱۲۹} و مورد پسند واقع شدن

این دو مورد تقریباً در همه‌ی شبکه‌های اجتماعی با نام‌های گوناگون وجود دارد. بازنشر توئیت در واقع امکانی است که به کاربر اجازه می‌دهد توئیتی که توسط فرد دیگری منتشر شده است را بازنشر دهد. در هنگام بازنشر، نام کسی که توئیت را منتشر کرده است، در توئیت بازنشر شده، درج می‌شود. این امکان در فیس‌بوک با نام "به‌اشتراک-گذاری"^{۱۳۰} وجود دارد.

امکان لایک کردن بدین معنا است که کاربران می‌توانند با استفاده از گزینه‌ای به همین نام که در توئیتر با علامت ♥ نشان داده می‌شود، علاقه‌ی خود را نسبت به یک توئیت نشان دهند.

۳-۱-۶- رابط کاربردی برنامه‌نویسی توئیتر

با توجه به رشد شبکه‌های اجتماعی و اهمیت داده‌های موجود در این سایت‌ها برای اهداف تحقیقاتی، تجاری، سیاسی، فرهنگی و ... ابزارهای دستیابی به این داده‌ها نیز فراهم شده است. ابزار موجود، رابط کاربردی برنامه‌نویسی نام دارد. اکثر شبکه‌های اجتماعی که امروزه وجود دارند برای کاربران خود، این ابزار را ارائه می‌دهند. از جمله‌ی

¹²⁰ Mention

¹²⁹ Retweet

¹³⁰ Share

این شبکه‌ها می‌توان به توئیتر، فیس‌بوک، لینکدین، اینستاگرام و ... اشاره کرد [۲۷]. در تحقیق حاضر نیز برای استخراج پایگاه‌داده از رابط کاربردی برنامه‌نویسی توئیتر استفاده شده است. برای استفاده از این ابزار، کافی است در شبکه‌ی توئیتر عضو شده و از طریق قسمت مربوط به توسعه‌دهندگان^{۱۳۱}، اپلیکیشن مورد نظر ثبت شود. پس از این مرحله، می‌توان از کتابخانه‌ی `twitter4j`^{۱۳۲} در زبان برنامه‌نویسی جاوا استفاده نموده و از امکانات رابط کاربردی برنامه‌نویسی توئیتر استفاده نمود. این شبکه‌ی اجتماعی محدودیت‌هایی را برای تعداد دفعاتی که می‌توان رابط کاربردی برنامه نویسی آن را صدا زد، گذاشته است. برای مثال اگر توسعه‌دهنده‌ای خواهان استخراج اطلاعات کاربران توئیتر باشد، هر ۱۵ دقیقه، فقط می‌تواند اطلاعات مربوط به ۱۸۰ کاربر را استخراج کند^{۱۳۳}. این محدودیت یکی از چالش‌های اساسی برای استخراج داده‌ها در پژوهش حاضر بوده است. برای توضیحات بیشتر در مورد پایگاه داده و نحوه‌ی استخراج داده‌ها از توئیتر به بخش ۳-۳ مراجعه نمایید.

۳-۲- ابزارهای مورد استفاده در پژوهش

این بخش به معرفی ابزارهایی که در تحقیق حاضر به کار رفته‌اند، اختصاص دارد. این ابزارها به صورت زیر می‌باشند:

۳-۲-۱- رابط کاربردی برنامه‌نویسی `Google Maps Geocoding`^{۱۳۴}

یکی از فاکتورهای به‌کاررفته شده در این تحقیق برای تعیین میزان تأثیرگذاری، محاسبه‌ی فاصله‌ی جغرافیایی بین شهر پاریس تا آن شهری است که توئیتر از آنجا ارسال شده است (برای اطلاعات بیشتر، به بخش ۳-۳ مراجعه فرمایید). از آنجایی که در یک توئیتر می‌توان محلی را که توئیتر از آنجا ارسال می‌شود، ذکر کرد، مسئله‌ی مهم به دست آوردن فاصله‌ی بین آن شهر تا پاریس در فرانسه است. برای محاسبه‌ی این فاصله، ابتدا باید طول و عرض جغرافیایی منطقه‌ی ارسال توئیتر به دست آید. استخراج این دو مقدار با استفاده از رابط کاربردی برنامه‌نویسی `Google Maps Geocoding` مقدور است. این رابط کاربردی برنامه‌نویسی با گرفتن آدرس محل جغرافیایی، می‌تواند با بررسی نقشه‌ی

^{۱۳۱} <https://dev.twitter.com/>

^{۱۳۲} <http://twitter4j.org/en/index.html>

^{۱۳۳} <https://dev.twitter.com/rest/public/rate-limits> [Feb. 5, 2016]

^{۱۳۴} <https://developers.google.com/maps/documentation/geocoding/intro> [Feb. 5, 2016]

گوگل، طول و عرض جغرافیایی آن را محاسبه کرده و به عنوان خروجی، ارسال کند. قالب درخواست ارسالی به ابزار ذکرشده، به صورت

<https://maps.googleapis.com/maps/api/geocode/output?parameter>

s

می‌باشد. خروجی Google Maps Geocoding رابط کاربردی برنامه‌نویسی می‌تواند JSON و یا XML باشد. برای مثال اگر طول و عرض جغرافیایی شهر شاهرود را بخواهیم، کافی است درخواست مورد نظر به صورت

<https://maps.googleapis.com/maps/api/geocode/json?address=Shahrood>

d

ارسال شود. خروجی JSON درخواست مورد نظر دارای دو زیرشاخه‌ی اصلی می‌باشد: status و result. قسمت اول، وضعیت درخواست را مشخص می‌کند که اگر پیغام “OK” برگردد، بیان‌گر پیدا شدن آدرس ارسالی در درخواست است. در قسمت دوم، زیرشاخه‌ی دیگری به نام geometry وجود دارد که اطلاعات جغرافیایی آدرس جست‌وجو شده را نمایش می‌دهد. یکی از قسمت‌های این بخش location است که در آن، عرض جغرافیایی با lat و طول جغرافیایی با lng مشخص می‌شوند. برای درخواست وارد شده، این مقادیر به ترتیب برابر با ۳۶,۴۰۶۲۲۴۱ و ۵۵,۰۱۶۲۶۸۹۹۹۹۹۹۹۹۹ خواهند بود. درخواست‌های ارسالی به رابط کاربردی برنامه‌نویسی Google Maps Geocoding با استفاده از زبان C# ارسال شده و جواب‌ها در پایگاه‌داده‌ی مورد نظر ذخیره شدند.

در استفاده از این رابط کاربردی برنامه‌نویسی، مانند بسیاری از رابط کاربردی برنامه‌نویسی‌های دیگر، محدودیت وجود دارد. محدودیت استفاده‌ی رایگان از رابط کاربردی برنامه‌نویسی Google Maps Geocoding، روزانه ۲۵۰۰ درخواست است که در هر ثانیه، تنها ۱۰ درخواست می‌تواند ارسال شود. برای درخواست‌های بیش‌تر از ۲۵۰۰ عدد، باید به ازای هر ۱۰۰۰ درخواست جدید، ۰,۵ دلار آمریکا پرداخت گردد^{۱۳۵}.

۳-۲-۲- نرم‌افزار Carrot

^{۱۳۵} <https://developers.google.com/maps/documentation/geocoding/usage-limits> [Feb. 5, 2016]

با توجه به نیازی که در این کار تحقیقاتی برای خوشه‌بندی^{۱۳۶} توئیت‌ها وجود دارد، از این نرم‌افزار استفاده شد. Carrot یک نرم‌افزار متن‌باز^{۱۳۷} می‌باشد که به زبان جاوا نوشته شده است. جدا از یک نرم‌افزار، کتابخانه‌ی آن نیز برای توسعه‌دهندگان و محققان در زبان‌های جاوا و سی‌شارپ موجود می‌باشد. کار با این نرم‌افزار بسیار راحت می‌باشد. رابط کاربری ساده و در حین حال قدرتمند آن، Carrot را به عنوان یکی از بهترین نرم‌افزارهای موجود برای خوشه‌بندی متون تبدیل کرده است.

این نرم‌افزار ورودی‌های مختلفی را برای متون در اختیار کاربران گذاشته است. از جمله‌ی آن‌ها می‌توان به قالب XML اشاره کرد. الگوی موجود برای قالب XML در شکل ۱-۳ آمده است.

```
<?xml version="1.0" encoding="UTF-8"?>
<searchresult>
  <query>Globe</query>
  <document id="0">
    <title>default</title>
    <url>http://www.globe.com.ph/</url>
    <snippet>
      Provides mobile communications (GSM) including
      GentXT, handyphones, wireline services, an
      broadband Internet services.
    </snippet>
  </document>
  <document id="1">
    <title>Skate Shoes by Globe | Time For Change</title>
    <url>http://www.globeshoes.com/</url>
    <snippet>
      Skaters, surfers, and showboarders
      designing in their own style.
    </snippet>
  </document>
  ...
</searchresult>
```

شکل ۱-۳ نمونه‌ای از ساختار ورودی فایل XML به نرم‌افزار Carrot^{۱۳۸}

در این قالب، هر document بیان‌گر یک متن می‌باشد. هر document، خود، می‌تواند شامل id، title، url و snippet باشد که توضیح هر یک به صورت زیر است:

- id: شماره‌ی سند مورد نظر

^{۱۳۶} Clustering

^{۱۳۷} Open source

^{۱۳۸} <http://download.carrot2.org/head/manual/index.html#figure.input-xml-format> [Feb. 5, 2016]

- title: عنوان سند
- url: اگر متن سند مورد نظر از یک سایت گرفته شده باشد، می توان در این قسمت، آدرس آن سایت را وارد نمود
- snippet: متن سند

برای گرفتن خروجی از این نرم افزار، تنها قالب XML وجود دارد. قالب خروجی Carrot در شکل ۲-۳ آمده است.

```

<?xml version="1.0" encoding="UTF-8"?>
<searchresult>
  <query>Globe</query>
  <document id="0">
    <title>default</title>
    <url>http://www.globe.com.ph/</url>
    <snippet>
      Provides mobile communications (GSM) including
      GenT&T, handyphones, wireline services, an
      broadband Internet services.
    </snippet>
  </document>
  <document id="1">
    <title>Skate Shoes by Globe | Time For Change</title>
    <url>http://www.globeshoes.com/</url>
    <snippet>
      Skaters, surfers, and snowboarders
      designing in their own style.
    </snippet>
  </document>
  ...
  <group id="0" size="60" score="1.0">
    <title>
      <phrase>com</phrase>
    </title>
    <group id="1" size="2" score="1.0">
      <title>
        <phrase>amazon.com</phrase>
      </title>
      <document refid="43"/>
      <document refid="77"/>
    </group>
    <group id="2" size="2" score="0.8">
      <title>
        <phrase>boston.com</phrase>
      </title>
      <document refid="4"/>
      <document refid="7"/>
    </group>
    ...
    <group id="7" size="48">
      <title>
        <phrase>Other Sites</phrase>
      </title>
      <attribute key="other-topics">
        <value type="java.lang.Boolean" value="true"/>
      </attribute>
      <document refid="1"/>
      <document refid="2"/>
      ...
    </group>
  </group>
  <group id="8" size="12" score="0.72">
    <title>
      <phrase>org</phrase>
    </title>
    <group id="9" size="2" score="1.0">
      <title>
        <phrase>en.wikipedia.org</phrase>
      </title>
      <document refid="9"/>
      <document refid="14"/>
      ...
    </group>
  </group>
  ...
</searchresult>

```

شکل ۲-۳ نمونه‌ای از ساختار خروجی فایل XML از نرم‌افزار Carrot^{۱۳۹}

در قالب خروجی، ابتدا فایل ورودی نمایش داده می‌شود و سپس هر خوشه در قالب یک group می‌آید. در هر group، موارد زیر از اهمیت ویژه‌ای برخوردارند:

^{۱۳۹} <http://download.carrot2.org/head/manual/index.html#section.architecture.output-xml> [Feb. 5, 2016]

- id: شماره‌ی خوشه
- size: تعداد اسنادی که در خوشه‌ی مورد نظر وجود دارند
- score: امتیازی که در رابطه با کیفیت هر خوشه داده می‌شود
- title: هر خوشه، خود شامل یک یا چند کلمه/عبارت کلیدی می‌شود که هر کدام در تگ^{۱۴۰} phrase جای می‌گیرد
- refid: شماره‌ی سندی می‌باشد که در قالب ورودی با id معرفی می‌شود.

الگوریتم‌های مختلفی برای خوشه‌بندی متون در Carrot پیاده‌سازی شده‌اند. از جمله‌ی این الگوریتم‌ها می‌توان به K-means، Lingo [۲۸] و Suffix Tree Clustering (STC) [۲۹] اشاره کرد. اشاره کرد. الگوریتم‌های Lingo و STC در هنگام خوشه‌بندی، ممکن است یک سند را در چند خوشه طبقه‌بندی کنند، حال آن‌که این اتفاق در K-means رخ نمی‌دهد. نکته‌ی دیگر در مورد این الگوریتم‌ها این است که Lingo و STC خوشه‌ای دارند که در آن اسنادی جای می‌گیرند که در خوشه‌های دیگر نمی‌توانند طبقه‌بندی شوند. این خوشه با نام Other Topics در Carrot نام‌گذاری شده است. این در حالی است که در K-means چنین خوشه‌ای وجود ندارد. ویژگی‌های هر یک از الگوریتم‌های مورد استفاده از Carrot در جدول ۳-۱ آمده است.

^{۱۴۰} Tag

جدول ۱-۳ مقایسه‌ی ویژگی‌های الگوریتم‌های *Lingo*، *K-means* و *STC* در نرم‌افزار *Carrot*^{۱۴۱}

ویژگی	Lingo	STC	K-means
تنوع خوشه‌ها	زیاد	کم	کم
برچسب خوشه‌ها	طول برچسب‌ها زیاد و غالباً توصیفی	طول برچسب‌ها کوتاه	برچسب‌های تک کلمه‌ای
مقیاس‌پذیری	پایین؛ برای بیش از ۱۰۰۰ سند، این الگوریتم زمان و حافظه-ی زیادی را مصرف می‌کند	بالا	پایین؛ همانند الگوریتم Lingo

نرم‌افزار *Carrot* قبل از عمل خوشه‌بندی، پیش‌پردازشی روی متون انجام می‌دهد.

این مرحله شامل مراحل زیر می‌باشد:

- ریشه‌یابی^{۱۴۲}: در این مرحله، کلمات به ساختار اصلی خود برمی‌گردند. به عنوان مثال، در زبان فارسی، کلمه‌ای مانند "انسان‌ها" به کلمه‌ی "انسان" که ریشه‌ی اصلی "انسان‌ها" است، برمی‌گردد. این نرم‌افزار با استفاده از ریشه‌یاب‌های *Snowball*، *Lucene* و *Morfologic* این عمل را به انجام می‌رساند.
- کلمات پرتکرار^{۱۴۳}: این قسمت مختص به حذف کلماتی است که در زبان‌های مختلف بسیار تکرار می‌شوند؛ مثلاً در زبان فارسی، کلماتی مانند "در"، "و"، "که" در این گروه جای می‌گیرند. برای پردازش یک متن، باید این کلمات حذف شوند. این کار توسط *Carrot* صورت می‌پذیرد.

^{۱۴۱} <http://download.carrot2.org/head/manual/index.html#section.advanced-topics.fine-tuning.choosing-algorithm> [Feb. 5, 2016]

^{۱۴۲} Stemming

^{۱۴۳} Stop words

■ برچسب‌های خاص^{۱۴۴}: یکی از امکانات مهم Carrot، فیلتر کردن برچسب‌هایی است که به هر خوشه اختصاص داده می‌شود. مثلاً اگر متونی که به عنوان ورودی به برنامه داده می‌شوند، همگی در زمینه‌ی ورزش باشند، می‌توان برای جلوگیری از برچسب‌گذاری‌های نامرتبط با موضوع ورزش، برخی از برچسب‌ها را در برنامه تنظیم کرد که آن‌ها در برچسب‌گذاری نیابند. مثلاً می‌توان برچسبی مانند "سیاست‌گذاری" را در Carrot تنظیم کرد که به عنوان برچسب در موضوع ورزشی ظاهر نشود. این برچسب‌ها به صورت عبارات باقاعده^{۱۴۵} در زبان جاوا تعریف می‌شوند. شکل ۳-۳ مثالی از این نوع عبارات باقاعده را نشان می‌دهد.

```
# stop label patterns for English
(?i)new
(?i)information (about|on).*
(?i)(index|list) of.*
(?i)(information|list|skip|join|cheap|access(es)?|corp(oration)?s?)
(?i).*(page|part|copyright) \d+.*
(?i)(official|offer(ing)?s?)|lists|uses?).*
(?i).*(known|information|offer(ing)?s?)|a range
```

شکل ۳-۳ نمونه‌ای از عبارات باقاعده‌ی تعریف شده در Carrot^{۱۴۶}

این نرم‌افزار از متون به زبان‌های گوناگون از جمله انگلیسی، عربی، فرانسوی، ترکی و ... پشتیبانی می‌کند.

۳-۲-۳ - کتاب‌خانه‌ی Document Similarity

یکی از قسمت‌های پایان‌نامه‌ی حاضر، بررسی شباهت بین توئیت‌ها است. برای این کار از کتاب‌خانه‌ی Document Similarity [۳۰] که در دانشگاه واترلو^{۱۴۷} کانادا تهیه شده است، استفاده شد. این کتاب‌خانه برای زبان سی‌شارپ فراهم شده است. نحوه‌ی

^{۱۴۴} Stop labels

^{۱۴۵} Regular expressions

^{۱۴۶} <http://download.carrot2.org/head/manual/index.html#section.lexical-resources.stoplabel-files> [Feb. 5, 2016]

^{۱۴۷} University of Waterloo

استفاده از Document Similarity بدین گونه است که ابتدا متون در قالب یک لیست در کتابخانه درج می‌شوند. پس از این کار، با فراخوانی متد مورد نظر، میزان شباهت هر دو متن به یکدیگر محاسبه می‌شود. برای محاسبه‌ی شباهت دو متن، از الگوریتم کسینوس^{۱۴۸} استفاده شده است. برای محاسبه‌ی میزان شباهت متون، ابتدا مراحل زیر انجام می‌شود:

- **پیش پردازش:** در این مرحله، تمام حروف در هر متن به حروف کوچک تبدیل و کلمات پرتکرار و اعداد حذف می‌شوند.
- **تبدیل متن به مدل فضای برداری^{۱۴۹}:** در این بخش، هر متن به یک بردار تبدیل می‌شود. بردارهای به وجود آمده از متون مختلف، ماتریس فضای برداری^{۱۵۰} را می‌سازند. هر درایه از ماتریس، وزن کلمه‌ی مشخصی را در یک متن نشان می‌دهد. هر وزن می‌تواند توسط الگوریتم‌های مختلفی مانند Binary Weighing، TF Weighing و TFIDF Weighing محاسبه شود.
- **محاسبه‌ی شباهت با استفاده از الگوریتم کسینوس:** برای محاسبه‌ی شباهت بین هر دو متن، بردارهای به دست آمده در مرحله‌ی قبل، دو به دو، به الگوریتم کسینوس داده می‌شود و نتیجه‌ی آن به برنامه برمی‌گردد.

عدد به دست آمده برای الگوریتم کسینوس در بازه‌ی [۰-۱] می‌باشد.

۳-۲-۴ - نرم‌افزار Gephi

از جمله مهم‌ترین نرم‌افزارها در حوزه‌ی بصری‌سازی شبکه‌های اجتماعی و به عبارت دقیق‌تر، گراف‌ها، می‌توان به Gephi اشاره کرد. این برنامه به زبان جاوا نوشته شده و متن‌باز می‌باشد. برای توسعه‌دهندگان نرم‌افزار و محققان نیز کتابخانه‌ی آن وجود دارد. واسط گرافیکی ساده، قدرت محاسباتی بالا، متن‌باز بودن و پیاده‌سازی مهم‌ترین

^{۱۴۸} Cosine algorithm

^{۱۴۹} Vector Space Model

^{۱۵۰} Vector Space Matrix

- رنگ کردن گره‌ها نسبت به فاکتورهای مختلف اندازه‌گیری مانند تعداد لینک‌های ورودی/خروجی
- تغییر اندازه‌ی گره‌ها نسبت به فاکتورهای مختلف اندازه‌گیری مانند تعداد لینک‌های ورودی/خروجی
- نمایش نام هر گره (اگر نام هر گره در فایل ورودی مورد نظر آمده باشد)
- ضخیم/نازک کردن لینک‌های بین گره‌ها
- تنظیم مکان گره‌ها/لینک‌ها با استفاده از الگوریتم‌های مختلف مانند Force Atlas، Yifun Hu، OpenOrd و ...
- نمایش خوشه‌های موجود در گراف با رنگ‌آمیزی‌های مختلف برای هر خوشه
- و ...

از جمله فاکتورهایی که Gephi توانایی اندازه‌گیری آن‌ها را دارد، می‌توان به موارد زیر اشاره کرد:

- محاسبه‌ی تعداد لینک‌های ورودی/خروجی برای هر گره
- محاسبه‌ی قطر شبکه^{۱۵۱}
- محاسبه‌ی چگالی شبکه^{۱۵۲}
- محاسبه‌ی سنج‌های مرکزیت
- محاسبه‌ی متوسط ضریب خوشه‌بندی^{۱۵۳}
- و ...

یکی دیگر از امکانات Gephi، قابلیت ذخیره‌ی شبکه‌ی به دست آمده در قالب png و pdf می‌باشد.

۳-۲-۵ - کتابخانه‌ی Twitter4j

همان‌طور که در بخش معرفی توئیتر توضیح داده شد، یکی از مهم‌ترین ویژگی‌های توئیتر، ابزار رابط کاربردی برنامه‌نویسی آن است. کتابخانه‌ی twitter4j که برای زبان جاوا طراحی

^{۱۵۱} Network diameter

^{۱۵۲} Network density

^{۱۵۳} Average clustering coefficient

شده است، این اجازه را به توسعه‌دهندگان و محققان می‌دهد که بتوانند از طریق آن درخواست‌های خود را به رابط کاربردی برنامه‌نویسی توییتر ارسال نمایند. این ابزار دارای مستندات قوی‌ای می‌باشد که کاربران می‌توانند با مراجعه به آن‌ها، نیازهای خود را پوشش دهند. با استفاده از twitter4j می‌توان لیست دنبال‌کنندگان/دنبال‌شوندگان، توئیتهای و ... را از توییتر استخراج نمود. برای این کار ابتدا باید هویت کاربر، با استفاده از کلیدهای امنیتی که توییتر پس از ثبت اپلیکیشن می‌دهد، بررسی شود. این کلیدها عبارت‌اند از: Consumer Key، Consumer Secret، Access Token و Access Token Secret.

۳-۳- پایگاه داده

۳-۳-۱- موضوع داده‌های جمع‌آوری شده

یکی از مهم‌ترین اتفاقاتی که در سال ۲۰۱۵ جهان را تحت تأثیر خود قرار داد، حملات متعدد گروه تروریستی داعش به نقاط مختلف دنیا و مخصوصاً فرانسه بوده است. اتفاقی که روز جمعه، ۱۳ نوامبر ۲۰۱۵ رخ داد، باعث شد نگاه مردم جهان این بار به سمت پایتخت پاریس معطوف شود. این حملات تروریستی در ۶ نقطه از شهر پاریس به وقوع پیوست^{۱۵۴}.

این اتفاق باعث شد کاربران زیادی در شبکه‌های اجتماعی مختلف، از جمله توییتر، مطالب بسیار زیادی بنویسند و هر یک از نگاه خود، این رخداد را تحلیل کنند. اهمیت این موضوع آن قدر بالا بود که توییتر، تأثیرگذارترین اتفاق سال ۲۰۱۵ را به حملات تروریستی پاریس اختصاص داد^{۱۵۵}. هشتک‌های مختلفی در توییتر اشاره به این رخداد داشتند که از آن جمله می‌توان #PrayForParis و #ParisAttacks را نام برد.

۳-۳-۲- نحوه‌ی جمع‌آوری داده‌ها

^{۱۵۴} <http://www.bloomberg.com/>

^{۱۵۵} <https://2015.twitter.com/most-influential> [Feb. 5, 2016]

همان‌طور که در بخش ۳-۱-۶ اشاره شد، رابط کاربردی برنامه‌نویسی توئیتر، ابزاری است که می‌توان داده‌های مورد نیاز را از این شبکه استخراج نمود. در این تحقیق نیز با استفاده از این ابزار، داده‌های مورد نظر استخراج شده‌اند.

از جمله مهم‌ترین چالش‌های موجود در این تحقیق، همان‌طور که در بخش ۱-۴ به آن اشاره شد، بحث فیلتر بودن شبکه‌ی اجتماعی توئیتر می‌باشد. برای غلبه بر این چالش، از سرور ابری ویندوز آژور^{۱۵۶} استفاده شد. برای اجرای برنامه‌ی مورد نیاز در رابطه با استخراج توئیتهای، ابتدا محیط توسعه‌ی نرم‌افزار^{۱۵۷} NetBeans و نرم‌افزارهای وابسته‌ی آن بر روی ویندوز آژور نصب شدند. پس از آماده‌سازی محیط، با استفاده از کتاب‌خانه‌ی twitter4j در زبان جاوا، کد برنامه نوشته و اجرا شد. از آنجایی که حادثه‌ی تروریستی پاریس در تاریخ ۱۳ نوامبر ۲۰۱۵ رخ داد، داده‌های مورد نظر از این تاریخ تا ۳ دسامبر ۲۰۱۵ جمع‌آوری شدند.

داده‌های جمع‌آوری شده در این تحقیق، دو نوع می‌باشند:

- داده‌های مربوط به توئیتهای: این داده‌ها با استفاده از هشتگ‌های #PrayForParis و #ParisAttacks جمع‌آوری شده‌اند.
- داده‌های مربوط به هر کاربر: پس از جمع‌آوری داده‌های مربوط به توئیتهای، داده‌های پروفایل هر کاربر نیز از طریق رابط کاربردی برنامه‌نویسی توئیتر جمع‌آوری شدند.

پس از جمع‌آوری داده‌ها، ۱،۷۹۲،۴۵۴ توئیتهای جمع‌آوری شدند. در قسمت دوم جمع‌آوری داده‌ها، اطلاعات ۵۹،۱۳۳ کاربر نیز استخراج شد. ذکر این نکته قابل توجه است که رابط کاربردی برنامه‌نویسی توئیتر با توجه به شخصی‌سازی بسیاری از پروفایل‌ها توسط کاربران این شبکه‌ی اجتماعی، امکان جمع‌آوری اطلاعات شخصی این افراد را ندارد.

با توجه به مدلی که در این پژوهش در نظر گرفته شده است، یکی دیگر از پارامترهایی که باید در پایگاه‌داده موجود باشد و از طریق رابط کاربردی برنامه‌نویسی توئیتر، قابل‌دسترس نیست، فاصله‌ی بین شهری است که توئیتهای آنجا ارسال شده تا شهر پاریس که مرکز حملات تروریستی بوده است. از آنجایی که شهری که توئیتهای

^{۱۵۶} Windows Azure Cloud Server

^{۱۵۷} Integrated Development Environment (IDE)

آنجا ارسال شده از طریق رابط کاربردی برنامه‌نویسی توئیتر قابل استخراج است، می‌توان شهر مورد نظر را به عنوان ورودی به رابط کاربردی برنامه‌نویسی Google Maps Geocoding داده و طول و عرض جغرافیایی شهر مورد نظر را به دست آورد. حال، با توجه به داشتن طول و عرض جغرافیایی پاریس و شهری که توئیتر از آنجا ارسال شده است، می‌توان فاصله‌ی آن‌ها را با استفاده از الگوریتم هاروه‌ساین^{۱۵۸} به دست آورد. با محاسبه‌ی این فاصله، مقدار آن در پایگاه‌داده‌ی مورد نظر برای هر توئیتر ذخیره شد.

۳-۳-۳ پارامترهای جمع‌آوری شده در داده‌ها

با توجه به دو قسمتی بودن فرآیند جمع‌آوری داده‌ها، پارامترهایی که در هر یک از قسمت‌ها استخراج شدند، در این بخش معرفی می‌شوند.

در بخش جمع‌آوری اطلاعات مرتبط با توئیترها، پارامترهای زیر با استفاده از رابط کاربردی برنامه‌نویسی توئیتر از هر توئیتر استخراج شدند:

- Username: نام کاربری افراد در توئیتر
- UserId: شناسه‌ی کاربر در توئیتر
- CreatedAt: تاریخ و ساعت انتشار توئیتر
- Message: متن توئیتر کاربر
- TweetLenght: طول توئیتر (بر اساس کاراکتر)
- FavoriteCount: تعداد دفعاتی که توئیتر مورد پسند کاربران واقع می‌شود
- RetweetCount: تعداد دفعاتی که توئیتر مورد نظر بازنشر داده می‌شود
- Language: زبانی که متن توئیتر با آن نوشته شده است
- Place: مکان جغرافیایی که توئیتر از آنجا ارسال شده است

^{۱۵۸} Haversine

شکل ۳-۵ نمونه‌ای از اطلاعات استخراج شده از توئیت کاربری با نام کاربری CarnegieEndow را با هشتگ #ParisAttacks نشان می‌دهد.

```
Username:
CarnegieEndow
UserId:
21332200
CreatedAt:
Fri Nov 27 23:15:05 UTC 2015
Message:
Following the #ParisAttacks, will France change its approach to Syria?
Pierre Vimont discusses: https://t.co/iFc5KG51w3
TweetLenght:
119
FavoriteCount:
5
RetweetCount:
1
Language:
en
Place:
null
```

شکل ۳-۵ نمونه‌ای از اطلاعات توئیت استخراج شده از کاربری با نام کاربری CarnegieEndow

در بخش اطلاعات مربوط به کاربرانی که توئیت‌های استخراج شده را منتشر کرده‌اند، ویژگی‌های زیر در نظر گرفته شدند:

- Username: نام کاربری افراد در توئیت
- UserId: شناسه‌ی کاربر در توئیت
- TweetCount: تعداد توئیت‌های ارسالی کاربر
- FollowingCount: تعداد افرادی که کاربر مورد نظر آن‌ها را دنبال می‌کند
- FollowersCount: تعداد افرادی که کاربر مورد نظر را دنبال می‌کنند
- FavouritesCount: تعداد دفعاتی که پست‌های دیگر کاربران مورد پسند کاربر مورد نظر واقع شده است

- **ListedCount**: تعداد لیست‌هایی که نام کاربر در آن‌ها قرار دارد و یا کاربر، خود، آن‌ها را ساخته است
- **Language**: زبانی که کاربر برای کار با توئیتر انتخاب کرده است
- **CreatedAt**: تاریخ و ساعتی که کاربر در توئیتر ثبت شده است
- **Description**: توضیحاتی که هر کاربر در مورد خود در پروفایلش دارد
- **Location**: مکان زندگی کاربر
- **MiniProfileImageUrl**: لینک آدرس عکس کوچک پروفایل کاربر
- **ProfileImageUrl**: لینک آدرس عکس پروفایل کاربر
- **TimeZone**: منطقه‌ی زمانی کاربر

شکل ۳-۶ نمونه‌ای از اطلاعات استخراج شده برای کاربر `joeeyrichardson` را نشان می‌دهد.

۳-۳-۴ - اطلاعات آماری پایگاه‌داده‌ی جمع‌آوری شده

در این بخش به بررسی اطلاعات آماری مرتبط با پایگاه‌داده‌ی جمع‌آوری شده از رابط کاربردی برنامه‌نویسی توئیتر می‌پردازیم. در این پایگاه داده، ۱,۷۹۲,۴۵۴ توئیت وجود دارد. از بین کاربرانی که توئیتهای را ارسال کرده‌اند، اطلاعات مربوط به ۵۹,۱۳۳ کاربر استخراج شدند. از بین توئیتهای جمع‌آوری شده، ۲۲,۵۲۰ توئیت هستند که کاربران، محلی که توئیت از آن‌جا منتشر شده است را ذکر کرده‌اند. بیش‌ترین تعداد بازنشر توئیت مربوط به توئیتی با نام کاربری `Louis_Tomlinson` می‌باشد که ۱۷۶,۰۲۱ بار بازنشر پیدا کرده است.

```
sername:
joeyrichardson
serId:
0433198
weetCount:
2831
ollowingCount:
62
ollowersCount:
43
avouritesCount:
145
istedCount:

anguage:
n-gb
reatedAt:
on Nov 16 16:57:35 UTC 2009
escription:
rench and Spanish student at University of Sunderland.
ocation:
outh Shields, South Tyneside
iniProfileImageURL:
http://pbs.twimg.com/profile_images/669824088815345665/KX00RLWY_mini.jpg
rofileImageURL:
http://pbs.twimg.com/profile_images/669824088815345665/KX00RLWY_normal.jpg
imeZone:
ondon
```

شکل ۳-۶ نمونه‌ای از اطلاعات استخراج شده از کاربری با نام کاربری *joeyrichardson*

توئیت کاربری با نام کاربری *katyperry* توانست با ۶۱,۰۶۹ بار مورد پسند واقع شدن، در این زمینه رکورددار باشد. بیش‌ترین تعداد دفعاتی که نام افراد در یک توئیت به کاررفته، مربوط به توئیت فردی با نام کاربری *whome51obama* می‌باشد که در شش توئیت خود، در هر یک، ۱۰ بار از نام افراد مختلف استفاده کرده است. بیش‌ترین تعداد هشتگ در یک توئیت نیز مربوط به کاربری با نام *PrayApp* می‌باشد که در یک توئیت، از ۱۵ هشتگ استفاده کرده است. جدول ۳-۲ و جدول ۳-۳ آمار تکمیلی در مورد پایگاه‌داده‌ی جمع‌آوری شده را به نمایش می‌گذارند.

جدول ۲-۳ اطلاعات آماری مربوط به پایگاه داده‌ی توئیت‌ها

عنوان	تعداد / میانگین
تعداد کل توئیت‌ها	۱,۷۹۲,۴۵۴
میانگین تعداد توئیت‌ها در هر روز	۸۹,۶۲۲
میانگین طول هر توئیت (بر حسب کاراکتر)	۱۱۳
میانگین تعداد مورد پسند واقع شدن هر پست	۰,۹۵
میانگین تعداد هشتگ‌های استفاده‌شده در هر توئیت	۰,۱۸
میانگین تعداد ذکر نام کاربران در هر توئیت	۰,۰۹
میانگین تعداد لینک آدرس استفاده‌شده در هر توئیت	۰,۰۵

جدول ۳-۳ اطلاعات آماری مربوط به پایگاه داده‌ی کاربران

عنوان	تعداد/میانگین
تعداد کل کاربران	۵۹,۱۳۳
میانگین تعداد توئیت‌های کاربران	۱۹,۳۷۸
میانگین تعداد دنبال‌کنندگان کاربران	۴,۲۹۹
میانگین تعداد کاربرانی که کاربر مورد نظر دنبال می‌کنند	۹۹۸
میانگین تعداد دفعاتی که پست‌های کاربران مورد پسند دیگران واقع شده‌اند	۵,۹۱۱
تعداد کاربرانی که مکان جغرافیایی خود را ذکر کرده‌اند	۴۲,۶۱۰
تعداد کاربرانی که توضیحاتی در پروفایل خود دارند	۵۹,۱۳۲
تعداد کاربرانی که زبان مورد نظر خود را در پروفایلشان ذکر کرده‌اند	۵۹,۱۳۳
تعداد کاربرانی که برای پروفایلشان عکس انتخاب کرده‌اند	۵۹,۱۳۳

۳-۴- بستر پیاده‌سازی

انجام این پروژه در دو بخش اصلی صورت گرفته است:

۱. جمع‌آوری داده

۲. پیاده‌سازی روش پیشنهادی

بخش جمع‌آوری داده، همان‌طور که در قسمت چالش‌های پروژه نیز بیان شد، به دلیل فیلتر بودن شبکه‌ی اجتماعی توئیتر، بر روی بستر ویندوز آژور صورت گرفت. سرور مورد استفاده دارای رم^{۱۵۹} ۱,۷۵ گیگابایت و حافظه‌ی ۱۳۵ گیگابایت بوده است. برای جمع‌آوری داده‌ها از زبان جاوا و محیط NetBeans 7.1 استفاده شد. کتاب‌خانه‌ی twitter4j ابزاری بود که به وسیله‌ی آن درخواست‌ها به رابط کاربردی برنامه‌نویسی توئیتر ارسال و نتایج آن در فایل ذخیره شد.

پس از جمع‌آوری داده‌ها، فایل‌ها برای پردازش به رایانه‌ی شخصی منتقل شده و پیاده‌سازی آن انجام شد. به منظور راحتی کار با داده‌های حجیم استخراج شده، اطلاعات موجود در فایل‌ها در پایگاه‌داده‌ی SQL Server 2014 ذخیره شدند.

پیاده‌سازی اصلی روش پیشنهادی با استفاده از زبان #.Net و محیط ویژوال استودیو و در چارچوب Net Framework 4.5 بوده است. مشخصات سیستمی که پیاده‌سازی روش در آن اجرا شده است به شرح زیر می‌باشد:

- نوع سیستم: Lenovo IdeaPad Z570
- پردازش‌گر: Intel® Core™ i5-2450M CPU @ 2.50GHz
- رم: 6 GB
- سیستم‌عامل: Windows 7 64-bit

۳-۵- پیش‌پردازش داده‌ها

۳-۵-۱- حذف هرزنامه‌نویسان^{۱۶۰}

یکی از دغدغه‌های مهم برای تشخیص افراد تأثیرگذار در یک شبکه‌ی اجتماعی، حضور و فعالیت هرزنامه‌نویسان در این حوزه بوده است [۱۶]. برای جلوگیری از تأثیر این کاربران در تعیین میزان تأثیرگذاری هر فرد در تحقیق موجود، با توجه به مدل ارائه شده در [۱۶]، آستانه‌ای برای

^{۱۵۹} RAM

^{۱۶۰} Spammers

تعداد توئیت‌های ارسالی این افراد در نظر گرفته شد. طبق پالایش در نظر گرفته شده، تنها افرادی در این تحقیق مورد بررسی قرار می‌گیرند که متوسط تعداد توئیت‌های آن‌ها در هر روز از زمان ثبت‌نامشان در توئیت‌ر تا روزی که داده‌ها جمع‌آوری شده است، کم‌تر از ۲ و بیش‌تر از ۰,۰۰۰۰۰۰۰۰۰۲ باشد. به عبارت دیگر، داریم:

$$(0.000000002) < \sum_{Date_{registered}}^{Date_{data_collection}} Days < (2) \quad (1-3)$$

با توجه به این معیار، تعداد توئیت‌ها و افراد مورد بررسی در این پژوهش، به ترتیب، به ۳۱،۸۱۵ و ۱۸،۶۷۷ کاهش یافتند.

۳-۵-۲- حذف هشتگ‌ها، نام کاربران و لینک‌های آدرس

قبل از مرحله‌ی شناسایی توئیت‌های مشابه با استفاده از الگوریتم کسینوس، توئیت‌ها پیش‌پردازش شدند. در این پیش‌پردازش، تمامی هشتگ‌ها، نام کاربران و لینک‌های آدرس در هر توئیت حذف شدند. این کار باعث می‌شود شباهت بین دو توئیت با دقت بیش‌تری صورت پذیرد. ضمن این‌که در کتاب‌خانه‌ی Document Similarity، یک مرحله‌ی دیگر پیش-پردازش صورت می‌پذیرد که طی آن، کلمات پر تکرار حذف می‌شوند.

۳-۵-۳- حذف توئیت‌ها با شباهت کم معنایی

از دیگر مراحل پیش‌پردازش داده‌ها زمانی است که روابط معنایی بین توئیت‌ها استخراج می‌شوند. در این مرحله از پیش‌پردازش، تنها روابطی انتخاب می‌شوند که درصد شباهت بین توئیت‌ها بیش از ۶۰ درصد باشد. این کار باعث می‌شود تعداد روابط معنایی بین توئیت‌ها به عدد ۱۶۱۹ برسد. جدول ۳-۴ آمار دقیق‌تری از پراکندگی تعداد روابط ارائه می‌دهد.

جدول ۳-۴ میزان پراکندگی روابط استخراجی از توئیت‌ها با شباهت بیش از ۶۰ درصد

تعداد روابط	بازه‌ی درصد شباهت
۴۰۵	۱۰۰ =
۴۲۸	[۹۰-۱۰۰)
۱۸۸	[۸۰-۹۰)

۲۸۳	[۷۰-۸۰)
۳۱۵	[۶۰-۷۰)

۳-۶- پارامترها

پارامترهای در نظر گرفته شده برای تعیین افراد تأثیرگذار در این پایان‌نامه به دو بخش زیر تقسیم می‌شوند:

- پارامترهای مرتبط با کاربر
- پارامترهای مرتبط با توئیت

پارامترهای مرتبط به کاربر، به دو بخش تقسیم می‌شوند: برخی از ویژگی‌ها نشأت گرفته از فعالیت کاربر در توئیت هستند و برخی دیگر ویژگی‌های موجود در پروفایل کاربر می‌باشند. این پارامترها به صورت زیر می‌باشند:

- پارامترهای نشأت گرفته از فعالیت کاربر
 - تعداد پست‌های کاربر (TweetCount)
 - تعداد دفعاتی که پست‌های دیگر کاربران مورد پسند کاربر مورد نظر قرار گرفته است (FavoriteCount)
 - تعداد دنبال‌کنندگان کاربر (FollowersCount)
 - تعداد کسانی که کاربر مورد نظر را دنبال می‌کند (FollowingsCount)
 - تعداد لیست‌هایی که کاربر دارد و یا نام وی در آن‌ها موجود است (ListedCount)
- پارامترهای موجود در پروفایل کاربر
 - (عدم) وجود لینک آدرس عکس کوچک پروفایل کاربر (MiniProfileImageUrl)
 - (عدم) وجود لینک آدرس عکس پروفایل کاربر (ProfileImageUrl)
 - (عدم) وجود توضیح کاربر در پروفایل خودش (Description)
 - (عدم) وجود زبانی که کاربر در پروفایل خود ذکر کرده است (Language)

○ (عدم) وجود مکان جغرافیایی که کاربر در پروفایل خود ذکر کرده است
(Location)

در هر توئیت، برخی از ویژگی‌ها توسط توئیت و برخی دیگر توسط دیگر کاربران مشخص می‌شود. با توجه به این دیدگاه، می‌توان پارامترهای در نظر گرفته شده برای هر توئیت را به صورت زیر عنوان کرد:

- پارامترهای مشخص شده توسط توئیت
 - تعداد هشتگ‌ها (HashtagCount)
 - تعداد دفعات ذکر نام دیگر کاربران (MentionCount)
 - تعداد لینک‌های استفاده شده (UrlCount)
 - تعداد کاراکترهای استفاده شده در توئیت (Length)
 - زمان انتشار توئیت (Time)
 - فاصله‌ی مکان جغرافیایی‌ای که توئیت از آنجا ارسال شده تا مکان مورد نظر که حادثه در آنجا رخ داده است (Distance)
- پارامترهای مشخص شده برای هر توئیت، نشأت گرفته شده از فعالیت دیگر کاربران
 - تعداد دفعات بازنشر توئیت (RetweetCount)
 - تعداد دفعات مورد پسند واقع شدن توئیت (FavoriteCount)
 - میزان شباهت توئیت‌های کاربران دیگر به توئیت مورد نظر (Similarity)

همه‌ی پارامترهای نام‌برده شده برای کاربران و نیز برخی از ویژگی‌های هر توئیت از طریق رابط کاربردی برنامه‌نویسی توئیت قابل استخراج می‌باشند. برخی از پارامترهای توئیت مانند تعداد هشتگ‌های استفاده شده را می‌توان از خودِ توئیت به دست آورد. در مورد پارامتری که توسط آن، میزان شباهت دیگر توئیت‌ها به توئیت مورد نظر مشخص می‌شود، باید بر روی توئیت‌ها تحلیل متن صورت پذیرد.

با توجه به ویژگی‌های ذکر شده، حال باید ابتدا این پارامترها تنظیم شده و سپس مدل مورد نظر ساخته شود.

۷-۳ - تنظیم پارامترها

۳-۷-۱- تعیین وزن پارامترها

برای ساخت مدل مورد نظر در تشخیص افراد تأثیرگذار در شبکه‌ی اجتماعی توئیت‌ر باید به هر یک از پارامترهای مورد نظر، وزنی اختصاص یابد. برای تعیین این اوزان از الگوریتم AHP [۱۷] استفاده شده است. این الگوریتم، یکی از روش‌های تصمیم‌گیری با معیارهای چندگانه سلسله‌مراتبی^{۱۶۱} است. در این الگوریتم، ابتدا ویژگی‌ها به صورت دو به دو مورد ارزیابی قرار می‌گیرند و اهمیت آن‌ها نسبت به هم توسط کاربر مشخص می‌شود و پس از انجام محاسبات، وزن هر یک از پارامترها به دست می‌آید. میزان اهمیت هر یک از ویژگی‌ها نسبت به دیگری در بازه‌ی [۱-۹] قرار می‌گیرد که ۱ بیان‌گر اهمیت یکسان بین دو ویژگی است و ۹ نشان‌دهنده‌ی بیش‌ترین میزان اهمیت یکی از ویژگی‌ها نسبت به دیگری است. جدول ۳-۵ و جدول ۳-۶ اهمیت دوبه‌دوی ویژگی‌های در نظر گرفته شده توسط نویسندگان را، به ترتیب، برای کاربران و توئیت‌ها نشان می‌دهد. خانه‌های سیاه رنگ، اهمیت در نظر گرفته شده را نشان می‌دهند.

جدول ۳-۵ نسبت اهمیت ویژگی‌های هر کاربر در الگوریتم AHP

ویژگی	نسبت اهمیت																ویژگی	
TweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	FollowingsCount
TweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	FollowersCount
TweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	FavoriteCount
TweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	ListedCount
TweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Language
TweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Description
TweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Location
TweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	MiniProfileImageUrl

^{۱۶۱} Hierarchical Multiple Criteria Decision Making

TweetCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr l
FollowingsCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	FollowersCoun t
FollowingsCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	FavoriteCount
FollowingsCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ListedCount
FollowingsCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Language
FollowingsCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Description
FollowingsCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Location
FollowingsCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	MiniProfileImage Url
FollowingsCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr l
FollowersCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	FavoriteCount
FollowersCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ListedCount
FollowersCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Language
FollowersCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Description
FollowersCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Location
FollowersCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	MiniProfileImage Url
FollowersCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr l
FavoriteCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ListedCount
FavoriteCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Language
FavoriteCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Description
FavoriteCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Location

FavoriteCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	MiniProfileImage Url
FavoriteCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr 1
ListedCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Language
ListedCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Description
ListedCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Location
ListedCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	MiniProfileImage Url
ListedCount	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr 1
Language	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Description
Language	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Location
Language	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	MiniProfileImage Url
Language	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr 1
Description	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Location
Description	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	MiniProfileImage Url
Description	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr 1
Location	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	MiniProfileImage Url
Location	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr 1
MiniProfileImage Url	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	ProfileImageUr 1

جدول ۳-۶ نسبت اهمیت ویژگی‌های هر توئیت در الگوریتم AHP

ویژگی	نسبت اهمیت																ویژگی	
Length	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	FovoriteCount
Length	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	RetweetCount
Length	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	HashtagCount
Length	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	MentionCount
Length	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	UrlCount
Length	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Distance
FovoriteCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	RetweetCount
FovoriteCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	HashtagCount
FovoriteCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	MentionCount
FovoriteCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	UrlCount
FovoriteCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Distance
RetweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	HashtagCount
RetweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	MentionCount
RetweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	UrlCount
RetweetCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Distance
HashtagCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	MentionCount
HashtagCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	UrlCount
HashtagCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Distance
MentionCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	UrlCount

MentionCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Distance
UrlCount	۹	۸	۷	۶	۵	۴	۳	۲	۱	۲	۳	۴	۵	۶	۷	۸	۹	Distance

پس از تنظیم میزان اهمیت دوبه‌دوی ویژگی‌های مرتبط با کاربران و توئیت‌ها، این الگوریتم شروع به محاسبه‌ی وزن هر یک از ویژگی‌ها خواهد کرد. وزن‌ها با استفاده از میانگین هندسی قابل اندازه‌گیری می‌باشند. پس از محاسبه‌ی میانگین هندسی برای هر یک از معیارها، اعداد به دست آمده، نرمال‌سازی می‌شوند. پس از این مرحله، اهمیت هر معیار به صورت نسبی برای هر یک از گزینه‌ها، که در این جا کاربران یا توئیت‌ها می‌باشند، مشخص می‌شود. ذکر این نکته قابل توجه است که اهمیت هر یک از معیارها در هر کدام از دو قسمتِ کاربران و توئیت‌ها، برای تمام پایگاه‌داده یکسان در نظر گرفته شده است. به طور مثال، میزان اهمیت ویژگی تعداد توئیت‌ها برای کاربر A بیش‌تر از کاربر B نمی‌باشد؛ بلکه این ویژگی برای همه‌ی کاربران با اهمیت یکسان در نظر گرفته شده است. نتیجه‌ی محاسبه‌ی اوزان پارامترهای موجود در جدول ۳-۷ و جدول ۳-۸ آمده است. آمده است. وزن هر یک از ویژگی‌ها در بازه‌ی [۰-۱] می‌باشد.

جدول ۳-۷- اوزان محاسبه شده برای پارامترهای هر کاربر توسط

الگوریتم AHP

ردیف	ویژگی	وزن محاسبه شده
۱	FollowersCount	۰,۴۰۵۰
۲	TweetCount	۰,۱۸۴۲
۳	ListedCount	۰,۱۷۵۰
۴	ProfileImageUrl	۰,۰۷۵۴
۵	Description	۰,۰۴۵۵
۶	MiniProfileImageUrl	۰,۰۳۷۷

۰,۰۲۵۸	Location	۷
۰,۰۲۳۱	FavoriteCount	۸
۰,۰۱۴۵	Language	۹
۰,۰۱۳۹	FollowingCount	۱۰

جدول ۳-۸ اوزان محاسبه شده برای پارامترهای هر توئیت توسط الگوریتم *AHP*

وزن محاسبه شده	ویژگی	ردیف
۰,۴۷۸۴	RetweetCount	۱
۰,۲۷۵۰	FavoriteCount	۲
۰,۱۱۰۴	Distance	۳
۰,۰۵۷۲	HashtagCount	۴
۰,۰۳۴۸	Length	۵
۰,۰۲۳۸	MentionCount	۶
۰,۰۲۰۴	UrlCount	۷

همان‌طور که مشخص است در پارامترهای مربوط به هر کاربر، پارامترهایی که نشأت گرفته از فعالیت وی هستند، نقش بسیار مهم و کلیدی ایفا می‌کنند. در سوی دیگر، پارامترهایی که در تعیین میزان شباهت توئیت‌ها به یکدیگر به کار می‌روند جزو پر اهمیت‌ترین ویژگی‌ها در رابطه با هر توئیت می‌باشند.

۳-۷-۲- تعیین میزان شباهت توئیت‌ها

این بخش به تعیین و تشخیص توئیت‌های مرتبط با توئیت مورد نظر اختصاص پیدا می‌کند. از آنجایی که حجم داده‌های جمع‌آوری شده برای توئیت‌ها زیاد می‌باشد و مشخص نمودن

یک رابطه‌ی معنایی بین یک توئیت با دیگر توئیت‌های پایگاه‌داده کار بسیار زمان‌بری است، لذا همان‌طور که در بخش پیش‌پردازش داده‌ها عنوان شد، برخی از توئیت‌ها از پایگاه‌داده حذف شدند. پس از این مرحله، برای بهبود عمل کرد برنامه در رابطه با ایجاد یک ارتباط معنایی بین توئیت‌ها، ابتدا این پیام‌ها با استفاده از الگوریتم STC در نرم‌افزار Carrot خوشه‌بندی شدند. برای این کار، پارامتر تعیین‌کننده‌ی تعداد خوشه‌ها برابر با ۵۰ در نظر گرفته شد. با توجه به این‌که این الگوریتم به طور پیش‌فرض یک خوشه به نام Other Topics در نظر می‌گیرد، تعداد خوشه‌ها به ۵۱ افزایش یافته. سپس، توسط الگوریتم STC، برای هر خوشه، عباراتی کلیدی استخراج و تخصیص داده شد. پس از خوشه‌بندی، همان‌طور که در بخش پیش‌پردازش اشاره شد، هشتگ‌ها، ذکر نام کاربران و نیز لینک آدرس‌ها در همه‌ی توئیت‌ها حذف شده و توئیت‌های هر خوشه با استفاده از الگوریتم کسینوس و توسط کتابخانه‌ی Document Similarity از نظر محتوایی مورد بررسی قرار گرفتند. برای تحلیل محتوایی، تنها روابطی انتخاب شدند که میزان شباهت بین دو توئیت را بیش از ۶۰ درصد نشان دهند. با توجه به این آستانه، ۱۶۱۹ رابطه بین توئیت‌ها به دست آمد که سهم هر خوشه به همراه تعداد توئیت‌ها و عبارات کلیدی آن در جدول ۳-۹ آمده است.

جدول ۳-۹ تعداد توئیت‌ها، عبارات کلیدی و تعداد روابط معنایی بین توئیت‌ها در هر خوشه با استفاده از الگوریتم STC در نرم‌افزار Carrot

شناسه‌ی خوشه	تعداد توئیت‌ها	عبارات کلیدی	تعداد روابط معنایی بین توئیت‌ها
۱	۷۶۲۵	British Embassy 33144513100	۲۸
		Irish Embassy 33144176700	
		Assistance in Paris	
۲	۳۵۷۱۵	Paris	۳۵۹
۳	۸۳۶۸	Concert Hall	۹
		Attackers Killed	

۱	French Media Says	۲۵۶۸	۴
	Police Operation at Concert		
۱	Number for those Concerned	۱۶۵۲	۵
	Foreign Office Number		
	Concerned about British Nationals		
۵۹	Attacks	۱۸۱۴۰	۶
۲۹	State of Emergency	۶۴۱۵	۷
۴	President Obama Says	۱۸۶۳	۸
	Outrageous Attempt to Terrorise		
۱۲۹	French	۱۶۳۵۹	۹
۵۷	People	۱۴۶۸۹	۱۰
۲	New York City Following	۱۲۲۹	۱۱
	Additional Police		
۳۹	Police	۱۲۶۶۹	۱۲
.	French President Francois Hollande	۱۱۹۸	۱۳
	Spoken of his Shock		
۳۷	France	۱۱۶۷۴	۱۴
۵۷	Closed	۷۰۷۸	۱۵
	Borders		
۵۱	Reports	۱۱۰۲۱	۱۶
۱۲	Hostages	۱۰۱۶۵	۱۷
۲۸	Terrorists	۹۰۶۹	۱۸
۲	Bataclan Concert Hall	۲۴۲۲	۱۹

۲	Stay Safe	۴۲۱۲	۲۰
۵	Killed	۸۰۹۶	۲۱
۲۶	Thoughts	۷۴۶۶	۲۲
۲۱	Terrorism	۷۲۰۷	۲۳
۱۵	Prayers	۷۲۰۲	۲۴
۳۳	Muslims	۷۱۲۷	۲۵
۰	11/9	۱۷۹۵	۲۶
	Carl		
	Stand with you Like		
۱۱	Bataclan	۶۹۶۸	۲۷
۶	Live	۶۸۱۱	۲۸
۳	Https	۶۷۵۷	۲۹
۰	Automatic Gunfire	۱۳۸۱	۳۰
	Inside Bataclan Concert Hall		
۱۰	Breaking	۵۹۱۶	۳۱
۰	Know so Far	۲۴۵۴	۳۲
۴	Know	۵۵۸۰	۳۳
۰	Learn	۱۱۵۴	۳۴
	Known Threat		
	City at this Time		
۱۰	Explosions	۵۵۵۵	۳۵

۲۷	President	۵۵۰۹	۳۶
۱	Police Officials	۱۰۶۴	۳۷
	Launched an Assault		
	Confirm Security Forces		
۴	Dead	۵۴۲۶	۳۸
۳	Declares State of Emergency	۱۱۱۱	۳۹
	Hollande Declares		
۵	Victims	۵۳۴۸	۴۰
۷	News	۵۱۵۵	۴۱
۵	Safe	۵۱۱۷	۴۲
۱	Tonight	۵۰۹۷	۴۳
۱۵	Pray	۵۰۵۶	۴۴
۲	Thoughts and Prayers	۲۵۹۸	۴۵
۷	Obama	۵۴۸۳	۴۶
۳	Happening	۴۳۰۰	۴۷
۰	French Police	۲۳۹۵	۴۸
۳	French President	۲۳۹۲	۴۹
۶	Inside	۴۱۹۰	۵۰
۴۸۰	Other Topics	۴۵۹۸۲	۵۱

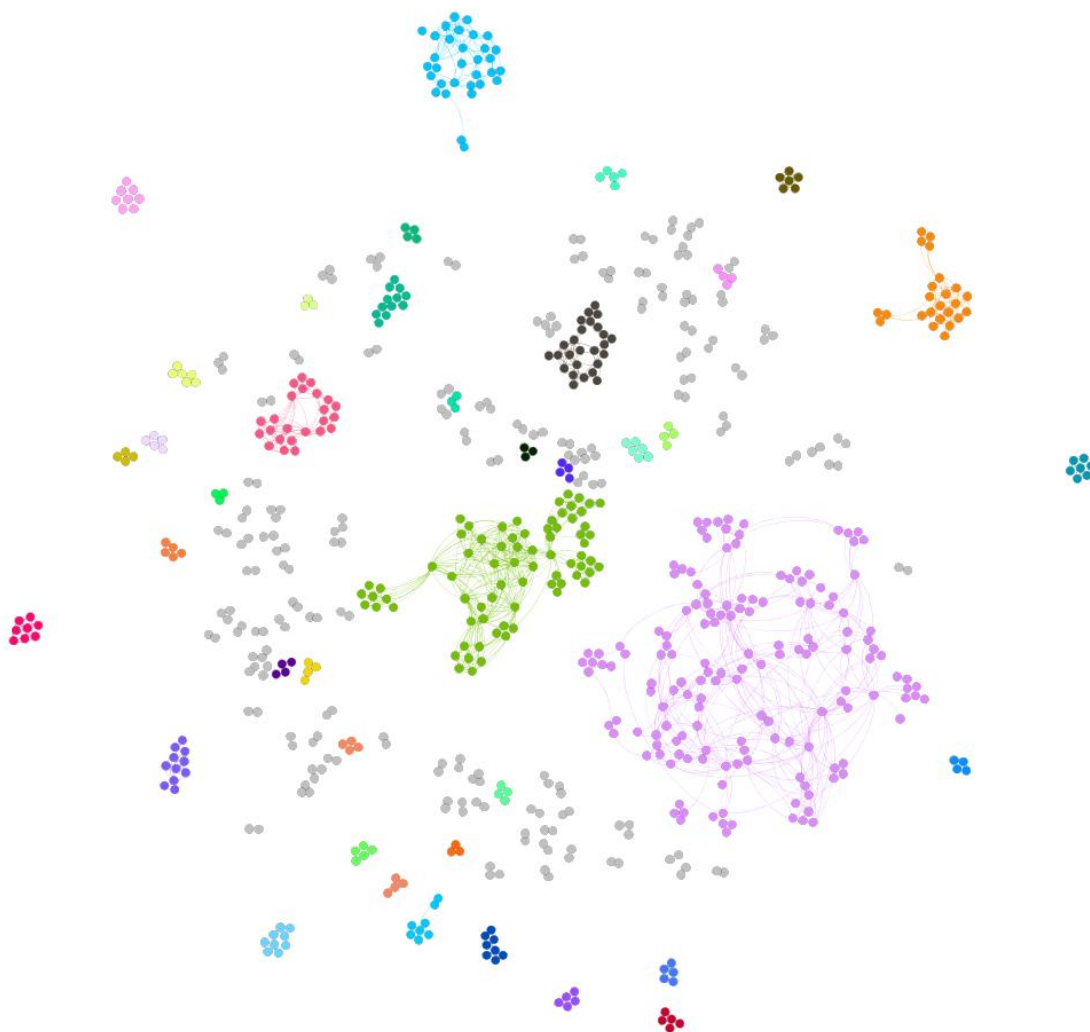
پس از به دست آوردن روابط موجود بین توئیت‌ها، این روابط توسط نرم‌افزار Gephi بصری‌سازی شدند. با توجه به این‌که الگوریتم STC برای خوشه‌بندی متون،

ممکن است یک توئیت را در چند خوشه جای دهد و از آن جایی که چنین امکانی برای نمایش داده‌ها در Gephi وجود ندارد، طبیعتاً از تعداد گره‌ها و لینک‌ها کاسته می‌شود. شکل ۷-۳ گراف ساخته شده توسط ارتباط معنایی بین توئیتهای پایگاه داده را در نرم‌افزار Gephi نشان می‌دهد.

جدول ۳-۱۰ آماری از گراف ساخته شده را نشان می‌دهد.

۳-۷-۳- تعیین فاصله بین محل انتشار توئیت و محل حادثه

یکی از پارامترهای در نظر گرفته شده در مدل ارائه شده، پارامتر فاصله تا محل حادثه می‌باشد.



شکل ۳-۷ نمایش بصری گراف ساخته شده از ارتباط معنایی بین توئیت‌ها

جدول ۳-۱۰ آمار به دست آمده از گراف استخراج شده از رابطه معنایی بین توئیت‌ها

مقدار اندازه‌گیری شده	معیار
۷۳۷	تعداد گره‌ها

۱۴۸۸	تعداد لینک‌ها
۴,۰۳۸	متوسط تعداد لینک‌های ورودی و خروجی
۲۷	بیش‌ترین تعداد لینک ورودی
۲۳	بیش‌ترین تعداد لینک خروجی
۳,۸۷۸	بیش‌ترین مقدار سنج‌های نزدیکی
۳۹۹,۱۳	بیش‌ترین مقدار سنج‌های بینابینی
۷	قطر شبکه
۲,۱۴۴	میانگین طول مسیر
۰,۰۰۳	چگالی شبکه

با توجه به ماهیت پایگاه‌داده‌ی در نظر گرفته شده در این تحقیق، وجود این پارامتر می‌تواند در تعیین افراد تأثیرگذار بسیار موثر باشد؛ زیرا افرادی که غالباً از نزدیک شاهد یک حادثه هستند، می‌توانند بهتر و دقیق‌تر در مورد آن نظر بدهند [۳]. همین نزدیکی فاصله باعث می‌شود گروه دیگری از افراد که از محل حادثه دور هستند، با تکیه بر نظرات شاهدان عینی، در مورد حادثه‌ی مورد نظر توثیق بنویسند. از آنجایی که حادثه‌ی تروریستی مورد مطالعه در شهر پاریس اتفاق افتاده است، لذا برای توثیق‌هایی که مکان انتشار آن‌ها ذکر شده است، این پارامتر مورد محاسبه قرار می‌گیرد. این کار با استفاده از رابط کاربردی برنامه‌نویسی Google Maps Geocoding و نیز الگوریتم هاروساین قابل اندازه‌گیری می‌باشد. پس از انجام عملیات پیش‌پردازش، ۵۰۴ توثیق بودند که محل انتشارشان در آن‌ها ذکر شده بود. جدول ۳-۱۱ آمار به دست آمده از بیش‌ترین و کم‌ترین فاصله‌ی بین مکان ارسال توثیق‌ها تا شهر پاریس نشان می‌دهد.

جدول ۳-۱۱ بیش‌ترین و کم‌ترین فاصله‌ی به دست آمده از توئیت ارسالی تا شهر پاریس

متغیر	فاصله (کیلومتر)	شهر/منطقه، کشور
بیش‌ترین فاصله	۱۶۹۵۹	سیدنی ^{۱۶۲} ، استرالیا
کم‌ترین فاصله	۸	سنت دنیس ^{۱۶۳} ، فرانسه

۳-۸- معرفی مدل

میزان تأثیرگذاری در توئیت می‌تواند از طرق مختلف اندازه‌گیری شود. مثلاً اگر به دنبال کاربری باشیم که دارای بیش‌ترین شهرت در بین دیگر افراد شبکه‌ی اجتماعی باشد، مشخصاً به دنبال تعداد دنبال‌کنندگان و یا فعالیت‌های وی، از قبیل تعداد توئیت‌ها، خواهیم رفت. اگرچه این معیارها می‌توانند نشان‌دهنده‌ی تعداد کسانی باشند که به پست‌های کاربر مورد نظر توجه دارند، اما نمی‌توانند بیان‌گر این باشند که آیا توئیت‌های وی واقعاً تأثیرگذار هستند یا خیر. به علاوه این‌که وقتی فردی در یک موضوع خاص به عنوان تأثیرگذار شناخته می‌شود، الزاماً به این معنی نیست که در دیگر موضوعات نیز می‌تواند تأثیرگذار باشد. در این پژوهش سعی شده است با توجه به پایگاه‌داده‌ی انتخابی، مدلی مناسب برای انتخاب فرد تأثیرگذار ارائه گردد.

مدل ارائه شده در تحقیق حاضر از دو بخش تشکیل می‌شود که یکی میزان تأثیرگذاری کاربران و دیگری میزان تأثیرگذاری توئیت‌ها می‌باشد. در این بخش به معرفی هر یک از این قسمت‌ها می‌پردازیم.

۳-۸-۱- مدل تأثیرگذاری کاربران

^{۱۶۲} Sydney

^{۱۶۳} Saint-Denis

با توجه به پارامترهای معرفی شده برای کاربران و نیز اوزان در نظر گرفته شده برای هر یک از آنها، مدل ارائه شده در رابطه‌ی (۲-۳) آمده است. از آنجایی که پارامترهای مربوط به پروفایل اشخاص به صورت کمی بیان نمی‌شوند، در این رابطه، تابع $E(T)$ تعریف شده است. مقدار این تابع زمانی که T وجود داشته باشد، برابر ۱ و در غیر این صورت برابر ۰ است. نکته‌ی قابل توجه، اثر منفی تعداد افرادی که کاربر مورد نظر، آنها را دنبال می‌کند. این معیار در مدل ارائه شده اثر منفی دارد و باعث می‌شود از میزان تأثیرگذاری کاسته شود، حال آن‌که دیگر معیارهای اندازه‌گیری به صورت مستقیم بر روی میزان تأثیرگذاری فرد، اثر مثبت دارند.

$$\begin{aligned} \text{Influence}_{\text{user}} = & w_{\text{FollowersCount}} * \# \text{FollowersCount} + w_{\text{TweetCount}} \\ & * \# \text{TweetCount} + w_{\text{ListedCount}} * \# \text{ListedCount} \\ & + w_{\text{ProfileImageUrl}} * E(\text{ProfileImageUrl}) + w_{\text{Description}} \\ & * E(\text{Description}) + w_{\text{MiniProfileImageUrl}} \\ & * E(\text{MiniProfileImageUrl}) + w_{\text{Location}} * E(\text{Location}) \\ & + w_{\text{Language}} * E(\text{Language}) - w_{\text{FollowingsCount}} \\ & * \# \text{FollowingsCount} - w_{\text{FavoritesCount}} * \# \text{FavoritesCount} \end{aligned}$$

(۲-۳)

۲-۸-۳- مدل تأثیرگذاری توئیت‌ها

این مدل نیز همانند مدلی که برای کاربران در نظر گرفته شد، از پارامترهای معرفی شده و اوزان آنها استفاده می‌کند. فرق عمده‌ی این دو مدل در تغییر میزان تأثیرگذاری بر حسب زمان می‌باشد. توئیت‌ها با افزایش زمان، میزان تأثیرگذاری خود را از دست می‌دهند. این اتفاق از آنجایی نشأت می‌گیرد که هر چه زمان جلوتر می‌رود میزان بازنشر و مورد پسند واقع شدن توئیت‌ها کم‌رنگ‌تر می‌شود. این اتفاق در مورد دیگر شبکه‌های اجتماعی مانند شبکه‌ی اجتماعی وبلاگ‌ها نیز صادق است. هر چه اختلاف بین زمانی که یک پست در وبلاگ ارسال شده (t_0) با زمانی که آن پست در حال بررسی است (t_1) افزایش پیدا کند، از میزان تأثیرگذاری آن پست کاسته می‌شود [۸]. با توجه به این موضوع، مدل ریاضی‌ای که در (۳-۳) آمده است، ارائه می‌گردد. در این رابطه، N تعداد روابطی است که در آنها، به توئیت مورد نظر ارجاع داده می‌شود (تعداد روابطی که در آنها، توئیت مورد نظر لینک ورودی دریافت می‌کند). $T_{\text{DataCollectionDate}}$ تاریخی است که در آن داده‌ها جمع‌آوری شدند و $T_{\text{TweetCreationDate}}$ تاریخی است که توئیت مورد نظر

در آن زمان منتشر شده است. اختلاف این دو تاریخ باعث کاهش میزان تاثیرگذاری یک توئیت خواهد شد.

$$\begin{aligned} \text{Influence}_{\text{tweet}} &= \left(w_{\text{HashtagCount}} * \#\text{HashtagCount} + w_{\text{Length}} * \#\text{Length} - w_{\text{Mention}} \right. \\ &\quad \left. * \#\text{Mention} - w_{\text{UrlCount}} * \#\text{UrlCount} + w_{\text{Distance}} * \frac{1}{\text{Distance}} \right) \\ &\quad + [(w_{\text{Retweet}} * \#\text{Retweet} + w_{\text{FavoriteCount}} * \#\text{FavoriteCount}) \\ &\quad * \left(\sum_{i \in 1}^N \text{Similarity}_i \right) / (T_{\text{DataCollectionDate}} - T_{\text{TweetCreationDate}})] \end{aligned}$$

(۳-۳)

از جمله موارد مرتبط با این مدل می‌توان به رابطه‌ی معکوس بین فاصله‌ی شهری که توئیت از آنجا ارسال شده تا شهر پاریس نام برد. به عبارت دیگر هر چه فاصله‌ی یک منطقه‌ی جغرافیایی از شهر پاریس بیش‌تر باشد، تأثیر آن توئیت کم‌تر می‌شود. ذکر این نکته قابل توجه است که انتخاب این پارامتر الزاماً نمی‌تواند در همه‌ی موارد مربوط به تشخیص افراد تأثیرگذار موثر باشد. این ویژگی زمانی تأثیر خود را نشان می‌دهد که ماهیت داده‌های جمع‌آوری شده این ویژگی را بطلبد.

۳-۸-۳ مدل جامع

با توجه به دو مدل ارائه شده در رابطه با کاربران و توئیت‌ها، حال می‌توان مدل جامع را در رابطه با این پژوهش ارائه داد. در این مدل، میزان تأثیرگذاری هر فرد با توجه به میزان تأثیر خود وی و توئیت‌هایش مشخص می‌شود. رابطه‌ی (۳-۴) این مدل را به صورت ریاضی تشریح می‌کند.

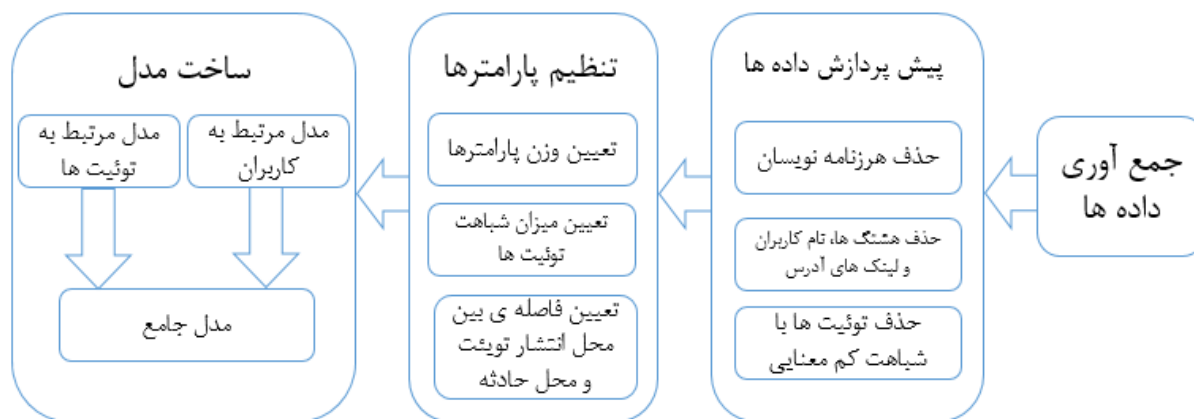
(۴-۳)

$$\text{Influence}_{\text{total}} = \text{Influence}_{\text{user}} * \sum_{i=1}^M \text{Influence}_{\text{tweet}_i}$$

در این مدل، M ، تعداد توئیت‌های منتشر شده از سوی کاربر مورد نظر می‌باشد.

۳-۹- نتیجه‌گیری

گام‌های برداشته شده از مرحله‌ی جمع‌آوری داده‌ها تا ساخت مدل مورد نظر، که در این فصل در نظر گذشت، به طور خلاصه در شکل (۳-۸) آمده است. همان‌طور که ملاحظه می‌کنید، پس از جمع‌آوری داده‌ها، پیش‌پردازش بر روی این داده‌ها انجام شد. در قسمت بعد، پارامترهای مدل، با استفاده از روش‌های توضیح داده شده در بخش‌های متناظر، تنظیم شدند. در نهایت، مدل‌های مرتبط با کاربران و توثیت‌ها، معرفی و سپس مدل جامع از روی آن‌ها ساخته شد. برای آگاهی از نقاط قوت مدل ارائه شده و نیز مقایسه‌ی آن با دیگر مدل‌های موجود، که در فصل ۲ از آن‌ها یاد شد، به بخش ۴-۵ مراجعه بفرمایید.



شکل ۳-۸ نمایش بصری گام‌های برداشته شده از جمع‌آوری داده‌ها تا ساخت مدل

۴- نتایج و تحلیل داده‌ها

۴-۱- امتیازهای مدل‌های تاثیرگذاری

۴-۱-۱- مدل تاثیرگذاری کاربران

با توجه به مدل ارائه شده برای کاربران، حساب‌های کاربری افراد حاضر در پایگاه داده مورد بررسی قرار گرفت. پس از انجام رتبه‌بندی، امتیازهای کسب شده، نرمال سازی شدند؛ بدین صورت که همه‌ی امتیازها بر بیش‌ترین امتیاز تقسیم شدند. با لحاظ کردن این موضوع، پنج نفر اول از لحاظ میزان تاثیرگذاری در جدول ۴-۱ مشخص شدند.

جدول ۴-۱- رتبه‌ی کاربران حاضر در پایگاه داده به همراه ویژگی‌های آن‌ها پس از اعمال مدل تاثیرگذاری کاربران

رتبه	نام کاربری	FollowersCount	TweetCount	ListedCount	ProfileImageUrl	Description	MiniProfileImageUrl	Location	FavouriteCount	Lang.	FollowingCount	امتیاز
۱	rssaha	۱۸۳۹۴۸۹	۷۵۸	۲۶	دارد	دارد	دارد	دارد	۶۰	دارد	۱۸۸	۱
۲	riotgames	۱۰۶۶۶۰۱	۲۹۴	۱۴۹۲	دارد	دارد	دارد	دارد	۱۸	دارد	۶	۰,۵۸
۳	YasirQadhi	۲۲۵۱۰۴	۲۰۳۲	۸۸۲	دارد	دارد	دارد	دارد	۰	دارد	۴۸	۰,۱۲
۴	AnonPress	۱۵۲۶۸۴	۱۹۲۱	۹۵۶	دارد	دارد	دارد	دارد	۱۶۱	دارد	۹۶	۰,۰۸۳
۵	respektor	۱۴۶۹۰۲	۹۷۲	۲۰۱۳	دارد	دارد	دارد	دارد	۱۸۹	دارد	۱۴۴	۰,۰۸۰

اگر چه ویژگی‌های منفی مدل، یعنی "تعداد توئیت‌هایی که مورد پسند کاربر مورد نظر واقع شده‌اند" (FavouriteCount) و "تعداد دنبال‌شوندگان توسط کاربر مورد نظر" (FollowingCount)، برای نفر اول، به ترتیب، ۳,۳ و ۳۱,۳ برابر همین ویژگی‌ها در نفر دوم است اما از آنجایی که وزن این دو ویژگی نسبت به وزن ویژگی‌های "تعداد دنبال‌کنندگان" (FollowersCount) و "تعداد توئیت‌ها" (TweetCount) کم‌تر است، همین عامل باعث سبقت گرفتن کاربر رتبه‌ی اول نسبت به نفر دوم شده است. دو ویژگی FollowingCount و TweetCount

برای کاربر **Rassaha**، به ترتیب، ۱,۷ و ۲,۵ برابر **Riotgames** است. هرچند ویژگی "تعداد لیست‌ها" (**ListedCount**) به صورت مثبت در مدل تاثیر می‌گذارد و نیز این ویژگی برای نفر دوم ۵۷,۳۸ برابر نفر اول است، اما با توجه به وزن به دست آمده برای آن در الگوریتم **AHP**، نتوانسته است تاثیر شدیدی در نزدیکی امتیاز نفر دوم به اولین نفر داشته باشد.

۴-۱-۲- مدل تاثیرگذاری توئیت‌ها

در این مرحله، توئیت‌های تاثیرگذار، پس از اعمال مدل ارائه شده برای آن‌ها و نرمال‌سازی امتیاز به دست آمده، استخراج شدند. نتیجه‌ی به دست آمده برای رتبه-بندی ۵ توئیت تاثیرگذار در جدول ۴-۲ آمده است. نکته‌ی قابل توجه در این توئیت‌ها، عدم ذکر نام کاربران و نیز تعداد صفر یا یک لینک آدرس در آن‌ها است که اثر منفی در مدل دارند. نکته‌ی دیگر عدم ذکر محلی است که توئیت از آنجا ارسال شده است. وجود این ویژگی در این توئیت‌ها می‌توانست بر میزان رتبه‌ی آن‌ها تاثیر بگذارد. با توجه به میزان ضریب به دست آمده برای هر یک ویژگی‌ها در الگوریتم **AHP**، همان‌طور که انتظار می‌رفت، تعداد بازنشر توئیت‌ها نقش بسزایی را در میزان تاثیرگذاری آن‌ها داشته است؛ به طوری که اولین توئیت در رتبه‌بندی کم‌تر از نصف توئیت دوم مورد پسند قرار گرفته و در ویژگی‌های دیگر نیز ضعیف بوده است، اما به واسطه‌ی حدود ۱,۵ برابر بودن تعداد بازنشر، توانسته در جایگاه اول قرار بگیرد. از دیگر نکات به دست آمده در پنج توئیت تاثیرگذار، می‌توان به غالب بودن هشتگ **#ParisAttacks** و عدم حضور هشتگ **#PrayForParis** اشاره کرد. با توجه به وقوع حادثه در تاریخ ۱۳ نوامبر ۲۰۱۵، همه‌ی پنج توئیت اول تاثیرگذار، در همین تاریخ و بلافاصله پس از حملات صورت گرفته است. از دو لینک استفاده شده در پنج توئیت تاثیرگذار، یکی از آن‌ها تصویری مربوط به تلفن سفارت‌خانه‌ی کشورهای مختلف در پاریس می‌باشد که در اولین توئیت آمده و لینک موجود در توئیت پنجم نیز عکسی از برج ایفل همراه با جمله‌ای برای هم‌دردی با مردم پاریس است.

از بین پنج توئیت اول، توئیتهای دوم و سوم به مخالفت مسلمانان جهان با این حملات اشاره کرده و با بازماندگان این حادثه ابرار هم‌دردی نموده‌اند. در توئیت دوم نیز قسمتی از سوره‌ی مبارکه‌ی مائده، آیه ۳۲، آمده است.

جدول ۴-۲- رتبه‌ی توئیت‌های حاضر در پایگاه‌داده به همراه ویژگی‌های آن‌ها پس از اعمال مدل تاثیرگذاری توئیت‌ها

رتبه	نام کاربری	متن توئیت	Rewteet Count	Favourite Count	Distance	Hashtag Count	Length	Mention Count	UrI Count	زمان انتشار توئیت	امتیاز
۱	parrysingh	Phone no. of major embassies in Paris. Pls share #PorteOuverte #ParisAttacks https://t.co/x31FW72QRA	۳۴۹۱	۶۸۹	-	۲	۱۰۰	۰	۱	۱۳,۱۱, ۲۰۱۵ - ۲۳:۵۲	۱
۲	malhilli	If you kill one human unjustly then ' you have killed all mankind' #Quran (5:32) #ParisAttacks Thoughts & prayers for victims + families	۱۹۸۳	۱۵۹۱	-	۲	۱۳۹	۰	۰	۱۳,۱۱, ۲۰۱۵ - ۲۳:۵۱	۰,۷۶
۳	OmarImranTweets	Before u blame the #ParisAttacks on 1.8billion Muslims just remember 99.9% of us are against this attack or any attack against humanity	۱۶۵۴	۱۰۷۲	-	۱	۱۳۷	۰	۰	۱۳,۱۱, ۲۰۱۵ - ۲۳:۵۶	۰,۶۰
۴	rssaha	I am completely shocked by the Paris blasts now , seems people have taken advantage of French hospitality. #ParisAttacks #ParisShooting	۱۳۸۸	۱۴۱۶	-	۲	۱۳۵	۰	۰	۱۳,۱۱, ۲۰۱۵ - ۲۳:۴۳	۰,۵۹
۵	KiingJaymes	Praying for the families of the victims that were killed and injured in the #ParisAttacks #Prayers4Paris https://t.co/xwfkQbcukQ	۱۴۵۱	۸۸۱	-	۲	۱۲۸	۰	۱	۱۳,۱۱, ۲۰۱۵ - ۲۳:۴۰	۰,۵۳

۳-۱-۴ - مدل جامع

پس از بررسی میزان تاثیرگذاری هر یک موارد کاربران و توئیت‌ها، حال نوبت به بررسی میزان تاثیرگذاری کلی در توئیت می‌باشد. طبق مدل ارائه شده، نتایج به صورت جدول ۳-۴ درآمده است.

جدول ۳-۴- رتبه‌ی کاربران بر اساس مدل جامع

رتبه	نام کاربری	امتیاز کاربری	جمع امتیاز توئیت‌ها	امتیاز کل
۱	rssaha	۱	۰,۵۹	۰,۵۹
۲	BBCkatyaadler	۰,۰۱۸	۴,۸۰۴	۰,۰۸۷
۳	OmarImranTweets	۰,۰۳۶	۰,۷۴	۰,۰۲۶
۴	nytgraphics	۰,۰۴۹	۰,۴۵	۰,۰۲۲
۵	respektor	۰,۰۸۰۵	۰,۲۵	۰,۰۲۰۴
۶	AnonPress	۰,۰۸۳	۰,۲۴	۰,۰۲۰۳
۷	YasirQadhi	۰,۱۲۳	۰,۱۲۶	۰,۰۱۵
۸	JosephineMcK	۰,۰۰۶	۱,۱۸	۰,۰۰۷
۹	yobynnad1127	۰,۰۱۲	۰,۴۶	۰,۰۰۵۷
۱۰	malhilli	۰,۰۰۴	۱,۱۵	۰,۰۰۵۴

در بین ده نفر اول رتبه‌بندی از لحاظ تاثیرگذاری، نام ۳ کاربر در بین پنج کاربر تاثیرگذار نیز دیده می‌شود که عبارتند از AnonPress، YasirQadhi و respektor و نیز نام دو نفر در بین کاربرانی دیده می‌شود که در مدل تاثیرگذاری توئیت‌ها جزو پنج نفر اول بودند:

malhilli و OmarImranTweets. در این بین، یک کاربر نیز هم در مدل تاثیرگذاری کاربران و هم در مدل تاثیرگذاری توئیت‌ها جزو پنج نفر اول بود که در مدل جامع، نفر اول لیست است: rssaha. در بین ده نفر تاثیرگذار در مدل جامع، تنها نفر اول است که میزان تاثیرگذاری کاربری‌اش از میزان تاثیرگذاری توئیت‌هایش بیش‌تر است. در ۹۰ درصد دیگر، این امر بر عکس است.

از نکات قابل توجه در نتایج به دست آمده می‌توان به تعداد توئیت‌های منتشر شده توسط هر کاربر در پایگاه‌داده‌ی موجود و رتبه‌ی آن‌ها اشاره کرد. جدول ۴-۴ این آمار را به نمایش می‌گذارد. همان‌طور که ملاحظه می‌شود، ۸۰ درصد کاربرانی که در بین ده نفر اول تاثیرگذارترین افراد هستند، کم‌تر از ۱۰ توئیت در مدت ۲۱ روز داشته‌اند. جدول ۴-۴ نشان می‌دهد میزان هم‌بستگی بین رتبه‌ی کاربران و میانگین امتیاز هر توئیت ۰,۶۱۹ می‌باشد که هم‌بستگی نسبتاً قوی‌ای است.

جدول ۴-۴- تعداد توئیت‌های کاربران در پایگاه‌داده و میانگین امتیاز هر توئیت

رتبه	نام کاربری	تعداد توئیت‌ها در پایگاه‌داده	میانگین امتیاز هر توئیت
۱	rssaha	۱	۰,۵۹
۲	BBCkatyaadler	۴۰	۰,۱۲۰
۳	OmarImranTweets	۲	۰,۳۷
۴	nytgraphics	۲	۰,۲۲۵
۵	respektor	۱	۰,۴۵
۶	AnonPress	۷	۰,۰۳
۷	YasirQadhi	۱	۰,۱۲۶
۸	JosephineMcK	۲۷	۰,۰۴
۹	yobynnad1127	۸	۰,۰۵

۰,۱۹	۶	malhilli	۱۰
------	---	----------	----

با توجه به مطالب بیان شده در دو پاراگراف قبل، می‌توان این‌گونه برداشت کرد که افراد، زمانی می‌توانند تاثیرگذار باشند که توئیت‌هایی با تاثیرگذاری بالا تولید کنند و در این بین، کیفیت توئیت‌ها مهم است (و نه کمیت آن‌ها).

۲-۴- روند تاثیرگذاری

پس از اعمال مدل ارائه شده در پایگاه‌داده‌ی موجود و نرمال‌سازی هر یک از امتیازات تاثیرگذاری مرتبط با توئیت‌ها و کاربران، روند تاثیرگذاری ده کاربری که در هر روز به عنوان تاثیرگذارترین افراد شناخته شدند، مورد بررسی قرار گرفت. پس از بررسی‌های صورت گرفته، تنها ۲۰ کاربر توانستند روند تاثیرگذاری خود را طی حداقل دو روز حفظ نمایند. این ۲۰ کاربر، روی هم رفته ۵۲ بار جزو ۱۰ کاربر تاثیرگذار در روزهای مورد بررسی بوده‌اند. در جدول ۴-۵ نام این کاربران به همراه تعداد دفعاتی که توانسته‌اند خود را در بین ده کاربر تاثیرگذار قرار دهند و نیز رتبه‌ی افراد در هر روز، آمده است.

همان‌طور که مشخص است کاربری با نام کاربری JosephineMcK توانسته است در چهار روز پیاپی، و در مجموع، هفت روز غیرپیاپی، روند حضور در بین ده نفر اول را از لحاظ تاثیرگذاری حفظ نماید و رتبه‌ی اول را به دست آورد. این کاربر گرچه در اولین دفعه‌ای که در بین افراد تاثیرگذار طبقه‌بندی شده است، رتبه‌ی چهارم را به خود اختصاص داده، اما در روزهای بعد روند خود را بهبود داده و در آخرین روزی که در بین ده کاربر تاثیرگذار بوده، نفر اول شده است.

جدول ۴-۵- رتبه‌ی کاربرانی که توانسته‌اند حداقل دو روز در بین ده فرد تاثیرگذار در هر روز باشند به همراه روزهای تاثیرگذاری آن‌ها

ردیف	نام کاربری	۱/۱۳	۱/۱۴	۱/۱۵	۱/۱۹	۱/۲۰	۱/۲۱	۱/۲۲	۱/۲۳	۱/۲۴	۱/۲۵	۱/۳۰	۱/۱	۱/۳
۱	JosephineMcK	۱	۱	۴	۳	۱	۲	۱	۱	۱	۱	۱	۱	۲

	۲	۱						۷			۲	M_Dorgan	۲	
					۳	۶	۵	۵				DonBishopSam	۳	
							۱		۱			BBCKatyaadler	۴	
								۶		۹	۴	Bedier	۵	
	۳			۱	۲							WeAreTheNobles	۶	
											۱	۵	Nytgraphics	۷
										۱		۷	AnonPress	۸
										۳		۱۰	yobynad1127	۹
							۱۰				۷		Rbsw	۱۰
									۴		۸		Changethechange	۱۱
						۱					۵		BishopAngaelos	۱۲
							۳				۸		CristinaSoler92	۱۳
										۲	۱۰		Theewesterner	۱۴
				۲						۹			Sophiepilgrim	۱۵
							۶	۲					JamesAALongman	۱۶
								۸	۴				Bbcbaxter	۱۷
						۴	۴						Andylines	۱۸
		۲			۴								KevStanf	۱۹
۱	۴												Jeriko_One	۲۰

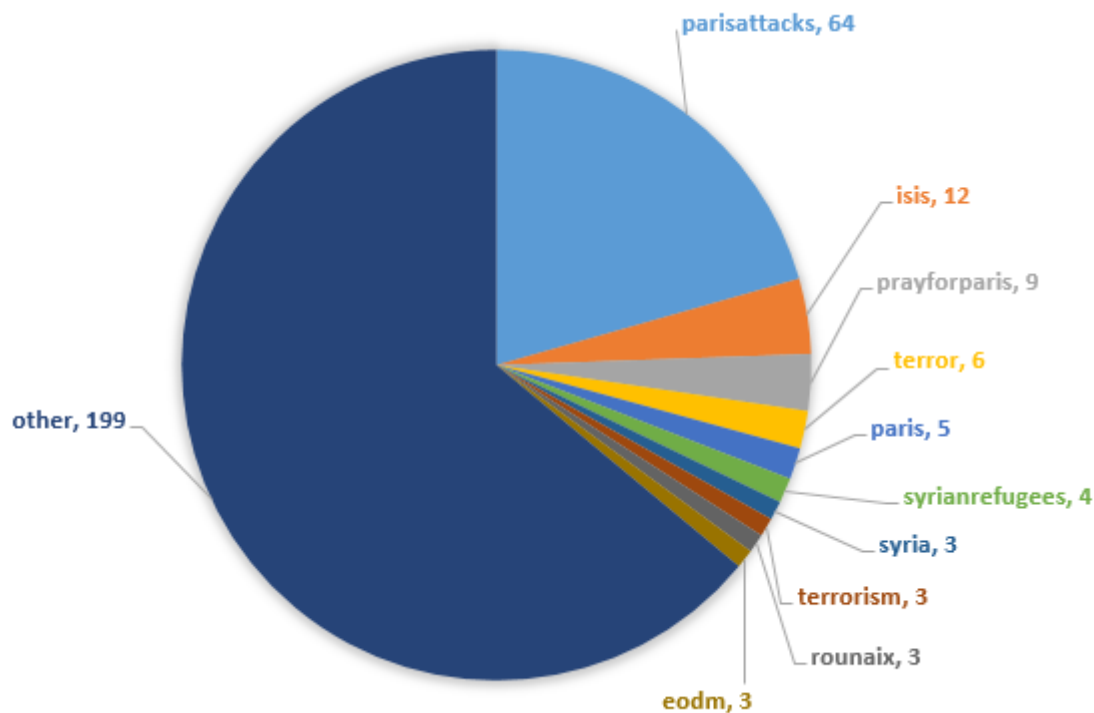
با توجه به جدول ۴-۵، در بازه‌ی ۲۱ روزه‌ای که داده‌ها جمع‌آوری شدند، تنها یک نفر توانسته است در ۳۳ درصد از این بازه، روند حضور در بین افراد تاثیرگذار را حفظ نماید. اگر تاریخ ۲۲ نوامبر و قبل از آن را نیمه‌ی اول و مابقی روزها را نیمه‌ی دوم بازه‌ی جمع‌آوری داده‌ها بگیریم، نفرات اول، دوم و سوم در جدول ۴-۵، توانسته‌اند خود را در هر دو نیمه‌ی بازه، تاثیرگذار نشان دهند. این موضوع برای دیگر افراد صادق نمی‌باشد؛ به طوری که آن‌ها نتوانسته‌اند در هر دو زیربازه قرار بگیرند. مثلاً کاربر BBCKatyaadler، همه‌ی روزهای تاثیرگذاری خود را در نیمه-ی اول بازه کسب کرده است.

با در نظر گرفتن این موضوع که ۱۷۸ نفر در بازه‌ی بین ۱۳ نوامبر تا ۳ دسامبر ۲۰۱۵ در بین ده کاربر تاثیرگذار قرار گرفته‌اند، ۱۵۸ نفر از آن‌ها تنها یک بار توانسته‌اند خود را در این بازه تاثیرگذار نشان دهند. به عبارت دیگر، حدود ۷۵ درصد افرادی که در هر روز، در بین افراد تاثیرگذار قرار می‌گیرند، نمی‌توانند روند تاثیرگذاری خود را، حتی برای یک روز دیگر، در موضوع حملات تروریستی پاریس، حفظ نمایند.

از دیگر نکات حائز اهمیت در این پژوهش را می‌توان این‌گونه بیان کرد که تنها ۴۱ درصد از رتبه‌های اول تا سوم مربوط به کسانی می‌شود که توانسته‌اند روند تاثیرگذاری خود را برای بیش از یک روز حفظ نمایند. به عبارت دیگر، ۵۹ درصد جایگاه‌های اول تا سوم به دست کاربرانی می‌افتد که تنها یک بار در دوره‌ی ۲۱ روزه، تاثیرگذار می‌شوند. این موضوع تعیین‌کننده‌ی صعود و سقوط ناگهانی ۵۹ درصد از افراد تاثیرگذار می‌باشد.

۴-۳- هشتگ‌های توئیت‌های تاثیرگذار

پس از بررسی پنج توئیتی اولی که در هر روز از بازه‌ی ۲۱ روزه تاثیرگذار تر از بقیه بودند، هشتگ‌های آن‌ها استخراج شد. هشتگ‌ها به حروف بزرگ و کوچک حساس هستند. مثلاً هشتگ‌های #prayforparis و #PrayForParis، دو هشتگ محسوب می‌شوند. برای جلوگیری از این حساسیت، ابتدا تمامی هشتگ‌ها به حروف کوچک تبدیل شدند. سپس تعداد این هشتگ‌ها در پنج توئیت برتر در هر روز مورد بررسی قرار گرفت. شکل ۴-۱ هشتگ‌های مورد استفاده را به همراه تعداد آن‌ها نشان می‌دهد. با توجه به نتایج به دست آمده، مشخص است که هشتگ #parisattacks بیش‌ترین تعداد تکرار را در پنج توئیت تاثیرگذار در هر روز داشته است. پس از آن هشتگ #isis که مربوط به گروه تروریستی داعش می‌شود، قرار دارد.



شکل ۴-۱ هشتگ‌های مورد استفاده در پنج توئیٹ تاثیرگذار در هر روز به همراه تعداد آن‌ها

۴-۴ هم‌بستگی بین متغیرهای موجود

جدول ۴-۶ و جدول ۴-۷ میزان هم‌بستگی بین ویژگی‌های در نظر گرفته شده به همراه نتایج به دست آمده به ترتیب برای کاربران و توئیٹ‌ها را نشان می‌دهد.

هم‌بستگی بین متغیرها از رابطه‌ی (۴-۱) قابل محاسبه است که در آن، X_i و Y_i مقادیر i -ام برای هر یک از متغیرها است. \bar{X} و \bar{Y} میانگین متغیرهای موجود هستند.

با توجه به نتایج به دست آمده، هم‌بستگی بین هر یک از ویژگی‌های در نظر گرفته شده برای کاربران و نتیجه‌ی تاثیرگذاری آن‌ها، به غیر از ویژگی‌های تعداد دنبال‌کنندگان و لیست‌ها، بسیار ضعیف می‌باشد. بیش‌ترین میزان هم‌بستگی مرتبط با ویژگی تعداد دنبال‌کنندگان با عدد ۰,۹۹ است. به جز هم‌بستگی کامل دو ویژگی عکس پروفایل و عکس کوچک پروفایل کاربران، هیچ هم‌بستگی قوی دیگری در بین پارامترهای مربوط به کاربران دیده نمی‌شود؛ به طوری که در رتبه‌ی دوم میزان هم‌بستگی‌ها، رابطه‌ی بین دو ویژگی تعداد لیست‌ها و دنبال‌کنندگان با عدد ۰,۲۸۸۲ است.

در رابطه با همبستگی‌های موجود بین ویژگی‌های در نظر گرفته شده برای توئیت‌ها و نیز عدد به دست آمده برای آن‌ها، باز هم شاهد همبستگی ضعیف بین پارامترها و میزان تاثیرگذاری هستیم به گونه‌ای که بیش‌ترین همبستگی مستقیم را تعداد دفعات مورد پسند واقع شدن پست‌ها با عدد ۰,۳۷۲۵ ایجاد کرده است. بیش‌ترین همبستگی معکوس نیز با ویژگی تعداد ذکر نام کاربران به دست آمده است (۰,۶۶۴۰۲-). در رابطه با همبستگی بین ویژگی‌ها نیز بیش‌ترین همبستگی را طول کاراکترهای توئیت و تعداد نام کاربران با عدد ۰,۳۷۲۷ به خود اختصاص دادند.

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (1-4)$$

جدول ۴-۶ - درجه‌ی هم‌بستگی ویژگی‌های در نظر گرفته شده و نتایج به دست آمده برای کاربران

	UserScore	FollowersCount	FavoriteCount	TweetCount	ListedCount	ProfileImageUrl	Description	MiniProfileImageUrl	Location	FollowingCount	Language
UserScore	۱	۰٫۹۹۹۵	۰٫۰۰۶۴	۰٫۰۴۴۹	۰٫۲۵۷۷	۰٫۰۰۰۲	-۰٫۰۰۰۰۴۳	۰٫۰۰۰۲	۰٫۰۲۱۵۵	۰٫۰۷۲۸۲	۰٫۰۲۰۱۱
FollowersCount	۰٫۹۹۹۵	۱	۰٫۰۰۲۴	۰٫۰۲۳۷	۰٫۲۸۸۲	۰٫۰۰۰۲	۰٫۰۰۰۲	۰٫۰۰۰۲	-۰٫۰۱۰۹	۰٫۱۲۵۵	۰٫۰۰۱۴
FavoriteCount	۰٫۰۰۶۴	۰٫۰۰۲۴	۱	۰٫۱۹۱۱	۰٫۰۰۳۳	۰٫۰۰۲۱	۰٫۰۰۲۱	-۰٫۰۳۵۶	-۰٫۰۰۹۹	۰٫۰۵۴۳	۰٫۰۰۰۸
TweetCount	۰٫۰۴۴۹	۰٫۰۲۳۷	۰٫۱۹۱۱	۱	۰٫۱۶۰۴	۰٫۰۰۱۰	۰٫۰۰۱۰	۰٫۰۰۱۳	۰٫۰۰۴۶	۰٫۱۳۷۶	-۰٫۰۰۰۴
ListedCount	۰٫۲۵۷۷	۰٫۲۸۸۲	۰٫۰۰۳۳	۰٫۱۶۰۴	۱	-۰٫۰۰۰۶	-۰٫۰۰۰۶	۰٫۰۰۱۲	۰٫۰۰۱۱	۰٫۰۸۲۸	۰٫۰۲۵۴
ProfileImageUrl	۰٫۰۰۰۲	۰٫۰۰۰۲	۰٫۰۰۲۱	۰٫۰۰۱۰	-۰٫۰۰۰۶	۱	-۰٫۰۰۰۰۳	۱	۰٫۰۰۸۵	۰٫۰۰۱۱	۰٫۰۰۹۸
Description	-۰٫۰۰۰۰۴۳	۰٫۰۰۰۲	۰٫۰۰۲۱	۰٫۰۰۱۰	-۰٫۰۰۰۶	-۰٫۰۰۰۰۳	۱	-۰٫۰۰۰۰۳	-۰٫۰۰۳۹	۰٫۰۰۱۵	۰٫۰۰۰۷
MiniProfileImageUrl	۰٫۰۰۰۲	۰٫۰۰۰۲	-۰٫۰۳۵۶	۰٫۰۰۱۳	۰٫۰۰۱۲	۱	-۰٫۰۰۰۰۳	۱	۰٫۰۰۸۵	۰٫۰۰۱۱	۰٫۰۱۵۴
Location	۰٫۰۲۱۵۵	-۰٫۰۱۰۹	-۰٫۰۰۹۹	۰٫۰۰۴۶	۰٫۰۰۱۱	۰٫۰۰۸۵	-۰٫۰۰۳۹	۰٫۰۰۸۵	۱	۰٫۰۰۶۵	۰٫۰۰۸۵
FollowingCount	۰٫۰۷۲۸۲	۰٫۱۲۵۵	۰٫۰۵۴۳	۰٫۱۳۷۶	۰٫۰۸۲۸	۰٫۰۰۱۱	۰٫۰۰۱۵	۰٫۰۰۱۱	۰٫۰۰۶۵	۱	-۰٫۰۵۴
Language	۰٫۰۲۰۱۱	۰٫۰۰۱۴	۰٫۰۰۰۸	-۰٫۰۰۰۴	۰٫۰۲۵۴	۰٫۰۰۹۸	۰٫۰۰۰۷	۰٫۰۱۵۴	۰٫۰۰۸۵	۰٫۰۵۴	۱

جدول ۴-۷- درجه‌ی هم‌بستگی ویژگی‌های در نظر گرفته شده و نتایج به دست آمده برای توئیت‌ها

	TweetScore	RetweetCount	FavoriteCount	Distance	HashtagCount	Length	MentionCount	UrlCount
TweetScore	۱	-۰,۱۵۱۳۷	۰,۳۷۲۵	۰,۰۰۵۶	۰,۱۸۸۰۹	-۰,۳۶۲۱۷	-۰,۶۶۴۰۲	-۰,۲۱۱۳۶
RetweetCount	-۰,۱۵۱۳۷	۱	-۰,۰۰۰۰۳	-۰,۰۳۵۸	-۰,۰۶۲۲	۰,۰۱۹۸	۰,۰۸۳۳	۰,۰۴۱۴
FavoriteCount	۰,۳۷۲۵	-۰,۰۰۰۰۳	۱	-۰,۰۱۳۳	۰,۰۰۲۱	-۰,۰۰۳۰	-۰,۰۴۷۷	-۰,۰۱۲۵
Distance	۰,۰۰۵۶	-۰,۰۳۵۸	-۰,۰۱۳۳	۱	-۰,۱۲۰۴	-۰,۰۵۵۱	۰,۰۸۶۱	۰,۰۴۸۸
HashtagCount	۰,۱۸۸۰۹	-۰,۰۶۲۲	۰,۰۰۲۱	-۰,۱۲۰۴	۱	۰,۰۸۸۶	-۰,۱۲۳۵	-۰,۰۱۸
Length	-۰,۳۶۲۱۷	۰,۰۱۹۸	-۰,۰۰۳۰	-۰,۰۵۵۱	۰,۰۸۸۶	۱	۰,۳۷۲۷	۰,۱۲۸۸
MentionCount	-۰,۶۶۴۰۲	۰,۰۸۳۳	-۰,۰۴۷۷	۰,۰۸۶۱	-۰,۱۲۳۵	۰,۳۷۲۷	۱	۰,۱۹۱۸
UrlCount	-۰,۲۱۱۳۶	۰,۰۴۱۴	-۰,۰۱۲۵	۰,۰۴۸۸	-۰,۰۱۸	۰,۱۲۸۸	۰,۱۹۱۸	۱

۴-۵- مقایسه‌ی روش پیشنهادی با دیگر مدل‌ها

در بسیاری از تحقیقاتی که در زمینه‌ی هوش مصنوعی انجام می‌شود، مانند تحقیقاتی که در حوزه‌ی پردازش تصویر صورت می‌پذیرد، معیار سنجش عمل کرد روش-های ارائه شده، می‌تواند وضوح یک تصویر یا میزان تشخیص کلمات در تصویر باشد. این معیارها دقیق و قابل اندازه‌گیری می‌باشند. در بررسی میزان تاثیرگذاری افراد در شبکه‌های اجتماعی، چنین معیاری وجود ندارد. به عبارت دیگر، بررسی کارآمد بودن مدل‌های ارائه شده در این حوزه، غالباً شخصی می‌باشد. مثلاً در [۹]، تعداد کاراکترهای به کار گرفته شده در یک پست به عنوان یکی از ملاک‌های تاثیرگذاری پست معرفی می‌شود، در حالی که این موضوع در [۱۰] رد می‌شود. با توجه به این موضوع، در این بخش، تفاوت‌ها و شباهت‌های مدل ارائه شده در پژوهش حاضر با برخی از مدل‌هایی که شرح آن‌ها در بخش "مروری بر کارهای پیشین" آمده است، به همراه نقاط ضعف و قوت هر یک، مورد بررسی قرار می‌گیرد.

۴-۵-۱- مقایسه با مدل‌های موجود در وبلاگ‌ها

در روش‌های ارائه شده برای تعیین وبلاگ‌های تاثیرگذار، با توجه به ماهیت پایگاه‌های داده و نیز مدل‌های ارائه شده در [۷]، [۸]، [۹] و [۱۰]، امکان بررسی ویژگی-های مربوط به وبلاگ‌نویس وجود ندارد. گرچه سرویس‌هایی مانند بلاگفا این قابلیت را برای وبلاگ‌نویسان فراهم کرده‌اند که اطلاعات خود را در پروفایلشان ذخیره کنند، اما این امکانات بسیار محدود بوده و تنها به مواردی مانند نام و نام خانوادگی وبلاگ‌نویس و علایق وی ختم می‌شود. این موضوع باعث می‌شود مدل‌های ارائه شده در این زمینه تنها به بررسی پست‌های موجود اکتفا کنند. در این بخش نیز، مدل ارائه شده در [۷]، بر خلاف روش پیشنهادی در پژوهش حاضر، زمان را در نظر نگرفته است؛ این در حالی است که زمان یکی از مهم‌ترین عوامل در تعیین میزان تاثیرگذاری در شبکه‌های اجتماعی می‌باشد. مدل‌های ارائه شده در [۸] و [۹]، گرچه زمان را در میزان تاثیرگذاری موثر دانسته و اعمال کرده‌اند، اما بر خلاف روش ارائه شده در این تحقیق، رابطه‌ی بین پست‌های وبلاگ‌ها را در نظر نگرفته و تنها به وجود لینک ورودی یا خروجی در هر وبلاگ بسنده کرده‌اند؛ این در حالی است که ممکن است وبلاگ‌نویسی مطالب خود را از وبلاگ دیگری بنویسد، اما هیچ لینک خروجی به پست‌های آن وبلاگ نداده باشد. مدل ارائه شده در [۱۰] که

تکمیل‌کننده‌ی روش‌های قبلی می‌باشد، تمامی موارد در نظر گرفته شده در مدل حاضر را در خود دارد؛ به علاوه این‌که از پارامترهای دیگری نیز مانند "خلاقیت پست" به عنوان یکی از ویژگی‌های افزایشنده‌ی تاثیرگذاری استفاده کرده است که در این پژوهش، با توجه به کم بودن طول هر توئیت، این ویژگی نمی‌توانست ملموس باشد.

۴-۵-۲- مقایسه با مدل‌های موجود در توئیت

در این بخش به بررسی و مقایسه‌ی چندین روش ارائه شده در توئیت برای تشخیص افراد تاثیرگذار می‌پردازیم.

مدل ارائه شده در [۳]، که پایگاه‌داده‌ی آن بی‌شبهت با پایگاه‌داده‌ی پژوهش حاضر نیست، به بررسی تعیین افراد تاثیرگذار در بهار عربی پرداخته است. در مدل ارائه شده، آن دسته از افرادی بالاترین رتبه‌ی تاثیرگذاری را به خود اختصاص داده‌اند که تنها دو ویژگی زیر را داشته باشند:

- توئیت آن‌ها دارای بیش‌ترین کلمات مرتبط با بهار عربی باشد
- مکانی که توئیت را از آنجا ارسال کرده‌اند به مناطقی که بهار عربی در آن‌ها اتفاق افتاده است، نزدیک باشد

این دو ویژگی از نظر نویسندگان مقاله، کافی هستند تا بتوان تاثیرگذارترین افراد را در چنین جریان‌هایی شناسایی نمود. حال آن‌که بر خلاف مدل ارائه شده در این پایان‌نامه، ویژگی‌های فردی کاربران از قبیل تعداد دنبال‌کنندگان و .. و نیز مهم‌ترین ویژگی‌های هر توئیت، از جمله تعداد بازنشرها و تعداد دفعات مورد پسند واقع شدن، مورد ارزیابی قرار نگرفته‌اند.

از دیگر تحقیقات صورت گرفته در این زمینه می‌توان به [۱۳] اشاره کرد. در این تحقیق از سه سنج‌ی توئیت، که ابتدایی نیز هستند، استفاده شده است. این سنج‌ها عبارتند از تعداد دنبال‌کنندگان، بازنشرها و نام کاربران. این در حالی است که بر خلاف مدل ارائه شده در این تحقیق، ویژگی‌های مهم دیگری از کاربران و توئیت‌ها، مانند تعداد دفعات مورد پسند واقع شدن توئیت و ...، مورد ارزیابی قرار نگرفته‌اند.

مدل ارائه شده در [۱۵]، تنها به بررسی محتوایی توئیت‌ها می‌پردازد. در این مدل، توئیت‌ها در طول زمان مورد ارزیابی قرار می‌گیرند و در هر نقطه که یک افزایش یا کاهش شدید نمره در محتوا وجود داشته باشد، تاثیرگذاری شناخته می‌شود. در

این مدل نیز، بسیاری دیگر از ویژگی‌هایی که در این پایان‌نامه در نظر گرفته شده، دیده نشده است.

یکی از کامل‌ترین مدل‌های ارائه شده در زمینه‌ی تعیین افراد تاثیرگذار در شبکه-ی اجتماعی توئیت، پایان‌نامه‌ی انجام شده در [۱۶] می‌باشد. بسیاری از فاکتورهای در نظر گرفته شده در پژوهش حاضر و پایان‌نامه‌ی مذکور مشابه می‌باشند. از جمله اختلاف‌های این مدل و روش ارائه شده در [۱۶]، می‌توان به موارد زیر اشاره کرد:

- عامل زمان، به عنوان پارامتری که نسبت معکوس با میزان تاثیرگذاری دارد، در [۱۶] دیده نشده است.
- رابطه‌ی بین توئیت‌ها در نظر گرفته نشده است و این پیام‌ها تنها از لحاظ تعداد بازنشر، مورد بررسی قرار گرفته‌اند.
- با توجه به اینکه، پارامتر "تعداد کسانی که کاربر مورد نظر آن‌ها را دنبال می‌کند" (FollowingCount) باعث کاهش میزان تاثیرگذاری کاربر می‌شود، این عامل در مدل ارائه شده در [۱۶] به صورت یک پارامتر مثبت در نظر گرفته شده است؛ حال آن‌که با توجه به این موضوع، این پارامتر در روش معرفی شده در این تحقیق، به صورت منفی مورد ارزیابی قرار می‌گیرد.
- نکته‌ی فوق در رابطه با تعداد پست‌هایی که مورد پسند کاربر مورد نظر واقع شده‌اند و نیز ذکر نام کاربران در توئیت نیز برقرار است.
- با توجه به ماهیت پایگاه‌داده‌ی جمع‌آوری شده در این تحقیق، در نظر گرفتن فاصله‌ی مکانی شهری که توئیت از آنجا ارسال شده است تا شهر پاریس، مهم می‌باشد. این پارامتر به دلیل بی‌اهمیت بودن در پایگاه‌داده-ی [۱۶]، در نظر گرفته نشده است.

جدول ۴-۸ رتبه‌بندی‌های به دست آمده توسط مدل ارائه شده در [۱۶] و مدل ارائه شده در این پژوهش را نشان می‌دهد. ذکر این نکته قابل توجه است که مدل ارائه شده در [۱۶]، تنها جمع امتیاز کاربر و هر توئیت را در نظر می‌گیرد، اما در جدول ۴-۸، ضرب امتیاز کاربر با مجموع امتیاز توئیت‌های وی در نظر گرفته شده است. با توجه به نتایج به دست آمده، ۵۰ درصد از ده نفر اول افراد تاثیرگذار طبق مدل ارائه شده در [۱۶]

همان رتبه‌ای را کسب کردند که در مدل ارائه شده در این تحقیق به دست آوردند. در بقیه‌ی موارد، غیر از مورد آخر، افراد نهایتاً ۳ پله با یکدیگر اختلاف داشتند.

جدول ۴-۸- مقایسه‌ی نتایج به دست آمده با اعمال مدل ارائه شده در [۱۶] و مدل ارائه شده در تحقیق حاضر

رتبه در مدل [۱۶]	نام کاربری	رتبه در مدل ارائه شده در تحقیق حاضر	امتیاز در مدل [۱۶]	امتیاز در مدل ارائه شده در تحقیق حاضر
۱	rssaha	۱	۰,۵۶۶۵	۰,۵۹
۲	BBCkatyaadler	۲	۰,۰۵۰۸	۰,۰۸۷
۳	OmarImranTweets	۳	۰,۰۲۴۶	۰,۰۲۶
۴	nytgraphics	۴	۰,۰۱۸۶	۰,۰۲۲
۵	respektor	۵	۰,۰۱۷۱	۰,۰۲۰۴
۶	YasirQadhi	۷	۰,۰۰۹	۰,۰۱۵
۷	AnonPress	۶	۰,۰۰۶	۰,۰۱۵
۸	KiingJaymes	۱۱	۰,۰۰۵	۰,۰۰۵۲
۹	malhilli	۱۰	۰,۰۰۴	۰,۰۰۵۴
۱۰	parrysingh	۱۳	۰,۰۰۳	۰,۰۰۴

۵- نتیجه‌گیری و کارهای آینده

در تحقیق حاضر، مدلی برای تشخیص افراد تاثیرگذار در شبکه‌ی اجتماعی توئیت ارائه گردید. جمع‌آوری داده‌ها با استفاده از رابط کاربردی برنامه‌نویسی توئیت و با استفاده از سرور Windows Azure صورت پذیرفت. موضوع داده‌های استخراج شده، حملات تروریستی پاریس در تاریخ ۱۳ نوامبر ۲۰۱۵ می‌باشد. این حادثه، طبق آمار توئیت، تاثیرگذارترین اتفاق در سال ۲۰۱۵ نیز می‌باشد.

برای ارائه‌ی مدل، ابتدا پارامترهای موجود مورد بررسی قرار گرفتند. پس از استخراج پارامترهای مورد نیاز، آن‌ها به دو گروه تقسیم شدند: (۱) پارامترهای مربوط به کابر و (۲) پارامترهای مربوط به توئیت. ویژگی‌های موجود در هر گروه، به الگوریتم AHP داده شده و با توجه به نسبت‌بندی اولیه‌ی پارامترها، وزن هر کدام محاسبه شد. در این مدل، توئیت‌ها از لحاظ محتوایی نیز مورد ارزیابی قرار می‌گیرند تا بتوان ارتباطی بین آن‌ها به دست آورد.

با توجه به مراحل گفته شده، مدل مورد نظر ساخته و میزان تاثیرگذاری هر یک از کاربران محاسبه شد.

با توجه به این موضوع که در زمینه‌ی تشخیص افراد تاثیرگذار در شبکه‌های اجتماعی، معیاری برای مقایسه‌ی روش‌ها وجود ندارد، مدل ارائه شده در این پژوهش با دیگر مدل‌های ارائه شده، به صورت کیفی، مورد ارزیابی قرار گرفت و نیز نتایج به دست آمده تحلیل شد.

از جمله کارهای آینده برای ادامه‌ی این پژوهش می‌تواند تحلیل محتوایی توئیت‌ها بر اساس احساسات باشد. این که نویسنده‌ی یک توئیت چه احساسی را در پیام خود جاسازی کرده است، می‌تواند باعث کاهش یا افزایش میزان تاثیرگذاری آن توئیت باشد. به طور مثال، توئیتی که با احساسات مثبت (شادی، ایثار و ...) نوشته شده است، می‌تواند در قیاس با توئیت‌های با احساسات منفی (غم، نفرت و ...) دارای تاثیرگذاری بیشتری باشند.

ترکیب مدل ارائه شده با دیگر روش‌های موجود در نظریه‌ی گراف، مانند سنجه‌های بینابینی و نزدیکی، از دیگر پیشنهادهایی است که برای ادامه‌ی کار توصیه می‌شود.

منابع

- [1] Hoang T., Cohen W. W., Lim E., Pierce D., Redlawsk D. P., (2013), “Politics , Sharing and Emotion in Microblogs”, ASONAM '13 Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 282-289, Niagara, Canada.
- [2] Overbey L. A, Paribello C., Jackson T., (2013), “Identifying Influential Twitter Users in the 2011 Egyptian Revolution”, **Social Computing, Behavioral-Cultural Modeling and Prediction: Lecture Notes in Computer Science**, 7812, pp. 377-385, 2013.
- [3] Kumar S., Morstatter F., Zafarani R., Liu H., (2013), “Whom Should I Follow? Identifying Relevant Users during Crises”, Proceeding HT '13 Proceedings of the 24th ACM Conference on Hypertext and Social Media, pp. 139-147.
- [4] Boyd D. M., Ellison N. B., (2008), “Social network sites: Definition, history, and scholarship”, **Journal of Computer-Mediated Communication**, 13, pp. 210–23.
- [5] Mogallapu A., (2011), Master of Science thesis, “Social Network Analysis of the Video Bloggers’ Community in Youtube”, Faculty of the Graduate School, Missouri University, United States of America.
- [6] Freeman L.C., (1978), “Centrality in social networks: conceptual clarification”, **Social Networks**, 1, pp. 215–239.
- [7] Agarwal N., Liu, H., Tang L., Yu P.S., (2008), “Identifying the influential bloggers in a community”, Proceedings of the international conference on Web search and web data mining - WSDM '08, pp. 207-217.
- [8] Akritidis L., Katsaros D., Bozanis P., (2009), “Identifying influential bloggers: Time does matter”, Proceedings - 2009 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2009, 1, pp. 76-83.
- [9] Akritidis L., Katsaros D., Bozanis P., (2011), “Identifying the productive and influential bloggers in a community”, IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 41, pp. 759-764.
- [10] Moh T. S., Shola S. P., (2013), “New factors for identifying influential bloggers”, Proceedings - 2013 IEEE International Conference on Big Data, pp. 18-27.
- [11] Xu B., Huang Y., Contractor N. S., Rd S., Tech D., (2013), “Structures of Broken Ties: Exploring Unfollow Behavior on Twitter”, Proceedings of the 2013 Conference on Computer Supported Cooperative Work, pp. 871-876.

- [12] Bollen J., Mao H., Zeng X., (2011), “Twitter mood predicts the stock market”, **Journal of Computational Science**, 2 (1), pp. 1-8.
- [13] Cha M., Haddadi H., Fabricio B., Gummadi P. K., (2010), “Measuring User Influence in Twitter: The Million Follower Fallacy”, Proc. International AAAI Conference on Weblogs and Social Media (ICWSM).
- [14] Bonacich P., Lloyd P., (2001), “Eigenvector-like measures of centrality for asymmetric relations”, **Social Networks**, 23, pp. 191–201.
- [15] Servi L., Elson S. B., (2014), “A Mathematical Approach to Gauging Influence by Identifying Shifts in the Emotions of Social Media Users”, **IEEE Transactions on Computational Social Systems**, 1 (4), pp. 180-190.
- [16] Metra I., (2014), Master of Science thesis, “Influence based exploration of Twitter Social Network”, Computer Science depart. Polytechnic University of Milan, Italy. Retrieved from <https://www.politesi.polimi.it/handle/10589/92709> [Jan. 5, 2016].
- [17] Thomas L. S., (1980), “**The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation**”, McGraw-Hill International Book Company, New York, United States of America.
- [18] Du Y., Gao C., Hu Y., Mahadevan S., Deng Y., (2014), “A new method of identifying influential nodes in complex networks based on TOPSIS”, **Physica A: Statistical Mechanics and its Applications**, 399, pp. 57-69.
- [19] Guimera R., Danon L., Guiler A. D., Giralt F., Arenas A., (2003), “Self-similar community structure in a network of human interactions”, **Physical Review**, 68 (6).
- [20] Girvan M., Newman M.E., (2002), “Community structure in social and biological networks”, **Proc. Natl. Acad. Sci.**, 99, pp. 7821–7826.
- [21] Chen D., Lü L., Shang M., Zhang Y., Zhou T., (2012), “Identifying influential nodes in complex networks”, **Physica A: Statistical Mechanics and its Applications**, 391, pp. 1777-1787.
- [22] Xie N., (2006), Master of Science dissertation, “Social network analysis of blogs”, University of Bristol.
- [23] Newman M. E. J., (2006), “Finding community structure in networks using the eigenvectors of matrices”, **Physical Review**, 74.
- [24] Leung C. K., Tanbeer S. K., Cameron J. J., (2014), “Interactive discovery of influential friends from social networks”, **Journal of Social Network Analysis and Mining**, 4 (1), pp. 154-166.
- [25] Han J., Pei J., Yin Y., (2000), “Mining frequent patterns without candidate generation”, **ACM SIGMOD**, pp. 1–12.

- [26] Campo-Ávila J. D., Moreno-Vergara N., Trella-López, M., (2013), “Bridging the Gap between the Least and the Most Influential Twitter Users”, **Journal of Procedia Computer Science**, 19, pp. 437-444.
- [27] Mathew A. R., (2013), “**Mining the Social Web**”, O’Reilly Media Inc., United States of America, 2.
- [28] Stanislaw O., Dawid W., (2005), “A Concept-Driven Algorithm for Clustering Search Results”, **IEEE Intelligent Systems**, 20, pp. 48-54.
- [29] Jerzy S., Dawid W., (2003), “Carrot2 and Language Properties in Web Search Results Clustering”, *Lecture Notes in Artificial Intelligence: Advances in Web Intelligence, Proceedings of the First International Atlantic Web Intelligence Conference*, pp. 240-249, Madrid, Spain.
- [30] PAMI Research Group, University of Waterloo, “Document Similarity” Internet: <http://pami.uwaterloo.ca/projects/lornet/software/docsimilarity.php>, 2006 [Jan. 5, 2016].
- [31] Song M., Kim M. C., Jeong Y. K., (2014), “Analyzing the political landscape of 2012 Korean presidential election in twitter”, **IEEE Intelligent Systems**, 29 (2), pp. 18-26.

Abstract

Due to the galloping growth of social networks and their usage as a crucial communication means in the world, their analysis is of a high importance. Identification of the most influential users, the people capable of changing other users' ideas and behavior in a virtual social network, is one of the most practical analyses in such networks. Such analysis is employed broadly in various fields including business, marketing and politics. The thesis is aiming at introducing a new method to identify the most influential users in Twitter social network.

A model is presented in this study to discover the influential Twitter users posting on Paris attacks in 2015. Such twitterers have been identified using the features regarding to tweets and the users posting them. The features are extracted employing Twitter Application Programming Interface (API). The proposed model divides the features into four groups: 1) the features in regards to users' interactions; 2) the features pointing to users' profiles; 3) the features shaped by the users reading the tweet; 4) the features included in the tweet. After preprocessing the collected data, the weights assigned to both user and tweet features are measured separately deploying AHP algorithm. As the other step toward identification of the most influential users, tweets are clustered and the semantic relationships between pairs of the tweets in each cluster are extracted. Then time factor, as one of the most effective features in changing the influence of the model, is employed. Next, the model is created and finally, it is applied to the database.

Results reveal the importance of tweets score compared to users score. The correlation between the rank of users and average score of tweets is measured as 0.619. The number uncovers the direct relationship between the quality of tweets and the rank of users. Results also show the quality of a single tweet message is more influential than the number of tweets. In other words, bringing an increase in the number of tweets does not guarantee for a growth in user's influence.

The trend of users' influence in a 21-day period discloses that 75 percent of the people ranked among the most influential users, cannot manage to hold their position for another day. It also shows 59 percent of the users ranked among the first 3 influential people lose their position as fast as they achieve it.

The hashtags used in the first 5 influential tweets of each day reveal a 20-percent allocation for #prayforparis. The next position goes to #isis with 3 percent.

The correlation between the features related to users and user scores show a strong relationship of 0.99 between the number of followers and user scores. The maximum correlation between tweet scores and related features is recorded 0.37 for the number of favourites.

The comparison between the proposed model and the other approaches toward identification of influential users reveals the comprehensiveness of the method introduced in this study. It covers the lacks of previous methods.

Keywords: Influential Twitter Users, Social Network Analysis, Recommendation System.



Shahrood University of Technology

Faculty of Computer Engineering and IT

**An Intelligent Recommendation System to Identify Influential
Users of a Social Network**

Behzad Rezaie

Supervisor:

Dr. Morteza Zahedi

Advisor:

Dr. Hoda Mashayekhi

Date: February 2016