




## رهیافتی برای انجام تحقیقات اجتماعی در توئیتر فارسی؛ مطالعه موردی انتخابات ریاست جمهوری ایران در سال ۱۳۹۶

دکتری علوم ارتباطات دانشگاه تهران، تهران، ایران.  حسین کرمانی \*

دانشجوی دکتری علوم ارتباطات دانشگاه علامه طباطبائی، تهران، ایران.  مرضیه ادهم

دانشجوی دکتری علوم ارتباطات دانشگاه تهران، تهران، ایران.  امیرعلی تفرشی

### چکیده

هدف این مقاله ارائه رویافتی جامع و دقیق برای انجام تحقیقات اجتماعی در توئیتر است. توئیتر در سال‌های اخیر به رسانه‌ای محبوب در بین کاربران تبدیل شده اما مسائل روش‌شناختی تحقیق در رسانه‌های اجتماعی در ایران به موازات گسترش استفاده از این رسانه‌ها توسعه نیافته است. این در حالی است که نبود رویافتی نظام‌مند می‌تواند به تشتت نظری و روشی و در نهایت بی‌اعتباری یافته‌ها در تحقیقات بر روی توئیتر فارسی بینجامد. ما در این تحقیق با ترکیب سه روش تحلیل شبکه اجتماعی، تحلیل محتوای مردم‌نگارانه و تحلیل گفتمان انتقادی رسانه‌های اجتماعی تلاش می‌کنیم این خلأ را رفع کنیم. رویافت ارائه شده شامل تمام مراحل تحقیق، از انتخاب بازه زمانی گردآوری داده‌ها تا تحلیل‌های آماری و نهایی است. این رویافت برخلاف رویافت‌های موجود، کاربر و محتوا (توئیتر) را به‌طور هم‌زمان در نظر گرفته و تحلیل می‌کند. همچنین راهکارهایی برای کدگذاری کاربرها و شناسایی ویژگی‌های جمعیت‌شناختی آن‌ها و توئیترها ارائه می‌کند که می‌تواند مبنایی برای انجام تحقیقات بعدی باشد.

**کلیدواژه‌ها:** توئیتر، رسانه‌های اجتماعی، تحلیل محتوا، تحلیل شبکه اجتماعی، تحلیل گفتمان انتقادی.

## مقدمه

اهمیت روش در انجام تحقیق عموماً کمتر از اهمیت نظریه در نظر گرفته می‌شود. درحالی‌که انتخاب یک روش مناسب، مبتنی بر اهداف تحقیق و هماهنگی با نظریه و دیگر اجزای پژوهش مانند انتخاب دقیق و بجای نظریه می‌تواند به محقق در رسیدن به یافته‌های مناسب‌تر و همچنین تحلیل‌های عمیق‌تر در راستای کشف و مشاهده موارد پنهان کمک کند. از طرف دیگر، مسائل روش‌شناختی تحقیق در رسانه‌های اجتماعی در ایران به موازات گسترش استفاده از این رسانه‌ها توسعه نیافته است و با خلای جدی در این حوزه مواجه هستیم. این تحقیق می‌کوشد تا با ارائه یک رهیافت روش‌شناختی جدید و ترکیبی برای تحقیق در توئیتر فارسی این خلأ را تا حد امکان رفع کند. این امر با توجه به اهمیت توئیتر برای کاربران ایرانی و همچنین نقش مؤثر آن در رویدادهای سیاسی-اجتماعی در سال‌های اخیر ضروری است.

توئیتر در بین رسانه‌های اجتماعی، مناسب‌ترین رسانه اجتماعی برای اطلاع‌رسانی‌های سریع و زنده بخصوص در رویدادهای سیاسی و اجتماعی در نظر گرفته می‌شود (هرمیدا، ۲۰۱۰؛ سدلر، ۲۰۱۸). برنارد (۲۰۱۶) معتقد است نقش تأثیرگذار توئیتر در حوزه سیاست، عمدتاً به قابلیت‌های فنی این رسانه اجتماعی برمی‌گردد، چیزی که پوئل و وندایک (۲۰۱۳) آن را «معماری تکنولوژیکی» این رسانه‌های می‌نامد. به همین دلیل، توئیتر در زمان‌هایی که بحران سیاسی‌ای در جامعه رخ می‌دهد (مانند بهار عربی، جنبش اشغال وال استریت، انتخابات ریاست جمهوری مولداوی در سال ۲۰۰۹ و اعتراضات به نشست G20 (پاپاچاریسی، ۲۰۱۴؛ امز، ۲۰۱۴) انتخاب اول کاربران بوده است.

کنش‌های کاربران توئیتر در موارد مختلف، اتفاقات سیاسی در ایران را نیز تحت تأثیر قرار داده است. انتخابات ریاست جمهوری سال ۱۳۸۸ و حوادث پس‌از آن یکی از مثال‌های بارزی است که تأثیرگذاری توئیتر بر روندهای سیاسی جامعه ایران را نشان می‌دهد تا جایی که برخی محققان از اعتراضات بعداز آن انتخابات با عنوان «انقلاب توئیتری» نام بردند

(مرزوف، ۲۰۰۹). نقش توئیتر در این رویداد سیاسی به گونه‌ای بود که مسئولان این رسانه اجتماعی به درخواست دولت آمریکا، زمان انجام تغییرات در توئیتر که موجب خارج شدن این سایت از دسترس کاربران ایرانی می‌شد را به تعویق انداختند (کارافانو، ۲۰۰۹). هرچند مناقشه آمیز بودن انتخابات آن سال، تحقیق درباره میزان دقیق این اثرگذاری را دشوار ساخته است اما موارد بعدی که بیانگر این اثرگذاری است ادعای ما را تأیید می‌کند. توئیتر در انتخابات مجلس شورای اسلامی در اسفند سال ۱۳۹۴ نیز اثرگذار بود. تجربه اثرگذاری توئیتر موجب شد تا در انتخابات شوراها، شهر و روستا و ریاست جمهوری در اردیبهشت ۱۳۹۶ نیز تمرکز کاندیداها بر این رسانه اجتماعی باشد (مرچنت و دیگران، ۲۰۱۸). علاوه بر این موارد، حضور چهره‌های سیاسی کشور در توئیتر نیز نشان‌دهنده اهمیت و اثرگذاری این رسانه است. در حال حاضر از رهبر و رئیس‌جمهور گرفته تا تعداد زیادی از وزرا و نمایندگان فعلی و سابق مجلس در توئیتر عضو هستند. این چهره‌های سیاسی در مواردی با استفاده از توئیتر توانسته‌اند بر روند اتفاقات اثرگذار باشند. یک مثال بارز در این زمینه، استفاده محمدجواد ظریف، وزیر خارجه در زمان انجام مذاکرات بر سر برجام است (هوشمند، ۲۰۱۵)؛ بنابراین هرچند تعداد کاربران ایرانی عضو در توئیتر به‌طور دقیق مشخص نیست، اما مثال‌های بالا و حضور چهره‌های سیاسی حکومت نشان‌دهنده اهمیت و اثرگذاری این رسانه اجتماعی است. اهمیت توئیتر موجب شده تا محققان توجه ویژه‌ای به استفاده از این شبکه میکرو بلاگ نویسی در تحقیقات اجتماعی داشته باشند.

### توئیتر؛ بستری برای تحقیقات اجتماعی

رسانه‌های اجتماعی با قابلیت‌های متعددی که دارند این نوید را داده‌اند که می‌توانند - و در حال حاضر تا حدودی توانسته‌اند - آغازگر عصری جدید در تحقیقات اجتماعی باشند (گولدر و میسی، ۲۰۱۲). در بین این رسانه‌ها، توئیتر مبنای اصلی تحقیقات متعددی بوده

است. یکی از دلایل این امر این است که توئیتر به‌عنوان «اندامواره مدل<sup>۱</sup>» برای کلان داده‌ها در نظر گرفته می‌شود (توفکی، ۲۰۱۴).

رسانه‌های اجتماعی، به‌طور خاص توئیتر، پارادیم تازه‌ای در گردآوری داده‌ها در تحقیقات اجتماعی را شکل داده‌اند (مک کورمیک و دیگران، ۲۰۱۵). با استفاده از توئیتر، محقق می‌تواند در زمان محدود و کم، حجم عظیمی از داده‌ها را دریافت کند. کاری که در روش‌های سنتی تحقیق در علوم اجتماعی مانند پیمایش تقریباً نشدنی است و در صورت امکان نیز به زمان، نیروی انسانی و هزینه زیاد نیاز دارد. همچنین استفاده از ابزار مناسب برای تحلیل این حجم داده‌ها، می‌تواند موجب آن شود که تحلیل‌ها در زمان بسیار کوتاه انجام شود. تبدیل و مناسب‌سازی داده‌های رسانه‌های اجتماعی برای تحلیل، نسبت به داده‌هایی که در پیمایش گردآوری می‌شود، بسیار راحت‌تر و سریع‌تر انجام می‌شود. همچنین داده‌های رسانه‌های اجتماعی، پنجره‌ای تازه در تحقیقات مبتنی بر مردم‌نگاری نیز گشوده‌اند. محقق می‌تواند کنش‌ها و کردارهای افراد را به‌طور زنده و بدون دست‌کاری و یا مبتلا شدن به خطر اعمال سلیقه شخصی، آن‌هم در حجم زیاد و به‌صورت لحظه‌به‌لحظه رصد و گردآوری کند. علاوه بر این موارد، داده‌های رسانه‌های اجتماعی پس از ثبت و گردآوری غیرقابل تغییر و ثابت هستند؛ بنابراین محقق نه‌تنها می‌تواند هر زمان که لازم بود به اصل داده‌ها دسترسی پیدا کند، بلکه می‌تواند این داده‌ها را بدون مواجه شدن با خطر تغییر و تحریف آن‌ها با دیگر افراد و محققان به اشتراک بگذارد. کرمانی (۱۳۹۵) قابلیت‌های رسانه‌های اجتماعی برای مردم‌نگاری را در سه طبقه پیش از ورود به میدان، در میدان و پس از میدان تشریح کرده است.

البته باید توجه داشت که رسانه‌های اجتماعی علیرغم این نقاط قوت، دارای نقاط ضعفی نیز برای استفاده در تحقیقات اجتماعی هستند. در وهله اول، استفاده از این داده‌ها و گردآوری آن‌ها به تخصص‌هایی در حوزه علوم کامپیوتر نیاز دارد که موجب می‌شود

---

1. Model organism

تعداد زیادی از محققان علوم اجتماعی نتوانند به خوبی از آن‌ها بهره ببرند (مک کورمیک و دیگران، ۲۰۱۵). البته از زمان مطرح شدن این مورد تاکنون، ابزارهای متعددی برای گردآوری، پالایش و تحلیل داده‌های رسانه‌های اجتماعی تولید شده که این مشکل را تا حد زیادی مرتفع کرده است. همچنین در پیمایش‌ها، با وجود اینکه محققان تعداد پاسخگویان کمتری دارند ولی کنترل بسیار بیشتر و تقریباً کاملی بر آن‌ها و شیوه پاسخگویی دارند. از این طریق، آن‌ها با احتمال بالاتر می‌توانند اطلاعاتی که می‌خواهند را به دست آورند. در تحقیق با استفاده از داده‌های رسانه‌های اجتماعی، محققان به حجم زیادی از داده‌ها دسترسی دارند که احتمالاً خیلی از آن‌ها اطلاعات مورد نیاز - مانند اطلاعات جمعیت شناختی - را به محققان نمی‌دهد. همچنین نگرانی‌هایی اساسی در بین محققان علوم اجتماعی در رابطه با نمونه‌گیری، تعمیم‌پذیری و کیفیت داده‌های رسانه‌های اجتماعی وجود دارد. به عنوان نمونه، مهت و شارکف (۲۰۱۳) بیان می‌کنند که امکان ندارد بتوان از داده‌های رسانه‌های اجتماعی آن‌گونه که در روش‌های سنتی نمونه‌گیری می‌شود، نمونه قابل اطمینان و تعمیم‌پذیری گرفت. یک مشکل دیگر که از مسائل بالا منتج می‌شود این است که محققان عمدتاً از کاربران و نه محتوا نمونه‌گیری کرده و تحقیق را بر مبنای چنین نمونه‌گیری‌ای بنا می‌کنند (مک کورمیک و دیگران، ۲۰۱۵). این رهیافت موجب می‌شود که تحقیقات عمدتاً صبغه محاسباتی و کمی داشته و عمق کافی و لازم را در حوزه علوم اجتماعی نداشته باشند.

با این وجود، محققان رهیافت‌های مختلفی برای تحقیق در توئیتر معرفی کرده‌اند (به عنوان مثال، رهیافت میدان شبکه‌ای توسط لینگل (۲۰۱۷)). نلسون (۲۰۱۶) نیز در تحقیق خود ۲۲۹ مقاله که از روش‌های کیفی برای تحقیق در رسانه‌های اجتماعی استفاده کرده‌اند را مرور کرده است. در بین رهیافت‌های معرفی شده، روش مک کورمیک و همکاران (۲۰۱۵) جالب توجه است اما ما در تحقیق حاضر می‌کوشیم نقاط ضعف این رهیافت را نیز برطرف کنیم. رهیافت مذکور تلاش می‌کند تا روشی برای گردآوری داده‌های جمعیت

شناختی در مورد کاربران به دست دهد. در وهله اول ما در این تحقیق بر اساس انتقاد مارشال (۲۰۱۲) از رویکردهایی که به تبدیل توئیت‌ها به داده صرف می‌پردازند و تأکید او بر روش‌های کیفی‌تر مبتنی برخوانش دقیق توئیت‌ها، از محدودیت تحقیق مک کورمیک و همکاران (۲۰۱۵) در تمرکز بر کاربران فراتر رفته و روشی را معرفی می‌کنیم که به‌طور هم‌زمان کاربران و محتوا را تحلیل می‌کند؛ بنابراین رهیافت ما، رهیافتی صرفاً محاسباتی نیست بلکه ترکیب روش‌های محاسباتی با تحلیل‌های متنی است. علاوه بر این، در رهیافت فوق محققان برای کدگذاری از بستر<sup>۱</sup> Amazon's Mechanical Turk استفاده کرده‌اند. از آنجا که این بستر تنها در ایالات متحده در حال حاضر قابل استفاده است، بنابراین برای تحقیقات روی توئیت در جوامع دیگر کاربرد ندارد؛ بنابراین ما رهیافتی را معرفی می‌کنیم که نیازی به این بستر نداشته باشد. در نهایت باید اشاره کنیم که این مقاله علاوه بر مشارکت علمی و روش‌شناختی در حوزه‌های فوق، تلاش می‌کند خلاء موجود در تحقیقات روی توئیت فارسی را نیز برطرف کند.

### تحقیقات توئیت در ایران

علیرغم اهمیت توئیت در ایران، تحقیقات انجام شده درباره نقش و اثرات استفاده از آن زیاد نیست. به‌طور خاص، روش‌شناسی مشخص و مناسبی برای انجام تحقیق در توئیت فارسی وجود ندارد. تحقیقات موجود درباره توئیت در ایران تا چند سال گذشته عمدتاً بر استفاده از این شبکه میکرو بلاگ نویسی در جریان انتخابات ریاست جمهوری در سال ۱۳۸۸ و اعتراضات پس از آن (جنبش سبز) متمرکز بود (به‌عنوان نمونه ن.ک: مغنی زاده، ۲۰۱۳؛ انصاری، ۲۰۱۲). در سال‌های اخیر توجه محققان به رویدادهای سیاسی دیگر مانند انتخابات ریاست جمهوری در سال ۱۳۹۲ نیز جلب شده و خصوصاً برخی تحقیقات با استفاده از روش تحلیل شبکه اجتماعی روی توئیت فارسی انجام شده است (مرچنت و

---

1. Platform

دیگران، ۲۰۱۶ و ۲۰۱۸؛ خضرای، ۲۰۱۹). این تحقیقات، هر چند، به درک ما از توئیتر فارسی عمق بیشتری بخشیده‌اند اما عمدتاً از لحاظ نظری قوی نبوده و همچنین روش‌شناسی بکار رفته در آن‌ها نظام‌مند و مشخص نیست به گونه‌ای که بتواند پایه‌ای برای تحقیقات بعدی فراهم سازد.

بنابراین، ما در این تحقیق می‌کوشیم تا یک رهیافت روش‌شناسی برای مطالعه توئیتر فارسی ارائه کنیم که در تحقیقات بعدی نیز بتواند استفاده شود. هر چند چنین روشی به‌طور عمومی برای تحقیق روی توئیتر می‌تواند استفاده شود منتها ما آن را روی توئیتر فارسی در انتخابات ریاست جمهوری سال ۱۳۹۶ اجرا کرده‌ایم و به همین دلیل ادعا می‌کنیم این رهیافت برای تحقیق روی توئیتر فارسی مناسب است.

انتخابات ریاست جمهوری سال ۱۳۹۶ برای انجام چنین تحقیقی، مورد مناسبی است. در واقع، مرجنت و دیگران (۲۰۱۸) معتقدند که پس از چند سال استفاده از رسانه‌های اجتماعی در کشور و همچنین استفاده از آن توسط گروه‌های سیاسی و کاندیداها در انتخابات مجلس شورای اسلامی سال ۱۳۹۴ و ریاست جمهوری سال ۱۳۹۲، در این انتخابات هم کاربران و هم احزاب به بلوغ نسبی در مصرف این رسانه‌ها و کنشگری در آن‌ها رسیدند. علاوه بر این، تحقیقات متعددی درباره نقش و چگونگی استفاده از توئیتر در رویدادهای سیاسی انجام شده است (به‌عنوان نمونه ن.ک: لارسون و موئه، ۲۰۱۲؛ المر، ۲۰۱۳؛ هوپکه، ۲۰۱۵؛ جکسون و فوکولت-ولز، ۲۰۱۵). با توجه به سنت موجود در استفاده از توئیتر برای تحقیق روی رویدادهای سیاسی و همچنین با توجه به خلأ تحقیقات در ایران و نقاط ضعف رهیافت‌های موجود؛ ما در این تحقیق می‌کوشیم روشی نظام‌مند و دقیق که هم‌زمان کاربران و محتوا را در توئیتر مطالعه کند ارائه دهیم. این تحقیق ما را قادر خواهد ساخت تا مشخص کنیم چگونه می‌توان از استفاده صرف از ابزارهای محاسباتی فراتر رفته و به تبیین‌های اجتماعی از توئیترها (بروکر و دیگران، ۲۰۱۶) - نه به‌عنوان داده برای روش‌های محاسباتی که به‌عنوان مواد معنادار برای تحقیقات اجتماعی- رسید. ما این

کار را با مطالعه کنشگری سیاسی در توئیتر فارسی با استفاده از این رهیافت روش‌شناختی انجام دادیم.

ما با تلفیق نظریه میدان بوردیو (۱۹۸۴) و مفاهیم همگان شبکه‌ای (بوید، ۲۰۱۰) و کردارهای شبکه‌ای (قالب‌بندی شبکه‌ای و روایتگری در توئیتر) ادعا کردیم که کنش‌های کاربران توئیتر در ایران (به‌عنوان اشکال جدید مشارکت سیاسی) به شکل‌گیری یک میدان سیاست تصویری<sup>۱</sup> در این رسانه اجتماعی انجامیده است. خوشه‌های<sup>۲</sup> مختلف کاربران، کنشگران شبکه‌ای این میدان هستند که از طریق اجرای کردارهای شبکه‌ای برای تصاحب قدرت در میدان تلاش می‌کنند. قالب‌بندی شبکه‌ای و تولید روایت‌ها، ازجمله این کردارهای شبکه‌ای هستند که ابزار مناسبی در اختیار کاربران برای منازعه بر سر قدرت قرار می‌دهد.<sup>۳</sup> البته تشریح و واکاوی این ادعا هدف این مقاله نیست و تنها در این مقاله به تشریح رهیافت روش‌شناختی پیشنهادی برای انجام این تحقیقات می‌پردازیم و به فراخور از مفاهیم فوق بهره خواهیم برد. در ادامه مقاله، مبانی نظری و همچنین مراحل عملیاتی این روش را به تفصیل توضیح می‌دهیم.

بستر نظری «رهیافت روش‌شناختی انجام تحقیقات اجتماعی در توئیتر فارسی» رهیافت روش‌شناختی ارائه‌شده در این تحقیق از ترکیب سه روش تحلیل شبکه اجتماعی، تحلیل محتوای مردم‌نگارانه (آلتهاید و اشنايدر، ۲۰۱۳) و تحلیل گفتمان انتقادی رسانه‌های اجتماعی (خسروی نیک، ۲۰۱۷) تشکیل شده است. ترکیب این سه روش با یکدیگر ما را قادر می‌سازد تا هم‌زمان هم کاربران و هم محتوای تولیدشده در توئیتر را مطالعه کنیم. علاوه بر این ما دو روش تحقیق نو را معرفی می‌کنیم که عمدتاً در ایران توجه خاصی به آن‌ها نشده است. تحلیل محتوا در ایران عمدتاً به تحلیل محتوای کمی و کیفی محدود می

1. Imaginary

2. Clusters

۳. نتایج این تحقیق در مورد انتخابات ریاست جمهوری به تفصیل در مقاله دیگری تشریح شده است.



شود و ما تحلیل محتوای مردم‌نگاری را به‌عنوان نوع خاصی از تحلیل محتوا که ابزار قدرتمندتری نسبت به دو رویکرد پیشین برای تحلیل متنی در اختیار محقق قرار می‌دهد معرفی می‌کنیم. همچنین تحلیل گفتمان انتقادی رسانه‌های اجتماعی روشی جدید برای تحلیل متنی در توئیت‌ها و دیگر رسانه‌های اجتماعی است که ابزاری به ما می‌دهد که دیگر روش‌های تحلیل گفتمان (انتقادی) فاقد آن هستند. استفاده از این دو روش با توجه به ویژگی‌های رسانه‌های اجتماعی منطقی‌تر از روش‌های دیگر است.

تحلیل شبکه اجتماعی: تحلیل شبکه اجتماعی یک روش نظام‌مند و محاسباتی برای استخراج داده‌های شبکه‌ای و تحلیل آن‌ها است. ما در این تحقیق ۴ مرحله اصلی در تحلیل شبکه اجتماعی را به‌صورت زیر معرفی می‌کنیم:

۱- گردآوری داده‌ها: در این مرحله با استفاده از ابزار مختلف داده‌های موردنیاز را گردآوری می‌کنیم. هرچند تحلیل شبکه اجتماعی قدمتی بیشتر از رسانه‌های اجتماعی دارد و گردآوری داده‌های موردنیاز برای چنین تحلیلی می‌تواند به‌صورت دستی انجام شود، لکن در حال حاضر این رسانه‌ها کار گردآوری داده‌های مربوط به خود را ساده کرده‌اند. در حال حاضر با استفاده از زبان‌های برنامه‌نویسی‌ای چون پایتون، آر و روبی می‌توان داده‌های موردنیاز را از رسانه‌های اجتماعی دریافت کرد. برخی ابزارهای رایگان نیز امکان گردآوری داده‌ها را بدون نیاز به داشتن مهارت‌های برنامه‌نویسی در اختیار محققان قرار می‌دهند (مانند نرم‌افزارهای Chorus، افزونه Ncapture و یا نرم‌افزار پولی NodeXL و وبسایت DiscoverText.com).

۲- پالایش داده‌ها: داده‌های خام دریافت شده از رسانه‌های اجتماعی اغلب به‌عنوان داده‌های درهم‌برهم<sup>۱</sup> و متشتت در نظر گرفته می‌شوند. این داده‌های خام دارای اطلاعات و قسمت‌های مختلفی هستند که برخی از آن‌ها موردنیاز نیستند. هنگامی که تعداد رکوردهای دریافتی، مثلاً تعداد توئیت‌ها- بسیار بالا باشد، حجم فایل‌های دریافت شده بزرگ خواهد بود. این فایل‌ها حاوی اطلاعاتی هستند که احیاناً در تحقیق کاربردی ندارند؛ بنابراین در

این مرحله می‌توان با استفاده از ابزار مناسب (با استفاده از زبان‌های برنامه‌نویسی یا استفاده از برخی ابزارهای رایگان مانند نرم‌افزار Tabeue) فایل‌هایی که تنها حاوی اطلاعات موردنیاز باشد ایجاد کرد. همچنین با توجه به هدف تحقیق، تمرکز پژوهشگران نه بر کل داده‌های دریافتی بلکه بر نوع خاصی از این داده‌ها خواهد بود. به‌عنوان مثال، در مورد توئیتر اغلب یکی از شبکه‌های منشن<sup>۱</sup>، ریتوئیت<sup>۲</sup> یا فالوشیپ<sup>۳</sup> موردتوجه قرار گرفته و صرفاً به آن داده‌ها نیاز است. در این مرحله این شبکه‌ها از دل شبکه اصلی اطلاعات استخراج می‌شوند.

۳- تحلیل داده‌ها: بعد از پالایش داده‌ها و استخراج شبکه موردنظر و با توجه به فرضیه‌ها و اهداف تحقیق، داده‌ها تحلیل می‌شوند. در این مرحله باید مفاهیم موردنیاز را بر اساس اهداف تحقیق مشخص کرد (مانند مفاهیم مربوط به مرکزیت<sup>۴</sup> یا بینیت<sup>۵</sup>) و ابزار مناسب برای محاسبه آن‌ها را بکار برد. در این مرحله نیز می‌توان با استفاده از زبان‌های برنامه‌نویسی مانند پایتون و آر، هر یک از سنجه‌های موردنیاز را محاسبه کرد. علاوه بر این زبان‌های برنامه‌نویسی، نرم‌افزارهایی مانند گفی<sup>۶</sup> امکان محاسبه این سنجه‌ها را بدون داشتن مهارت‌های برنامه‌نویسی فراهم کرده‌اند.

۴- بصری سازی داده‌ها: مرحله چهارم، یعنی بصری سازی داده‌ها عمدتاً برای ارائه مطلوب‌تر نتایج و به دست دادن فهمی تصویری از شبکه و روابط آن صورت می‌گیرد. در این مرحله، نتایج به دست آمده از تحلیل‌ها را می‌توان به صورت بصری و در قالب گراف‌های خاصی نمایش داد. به‌عنوان مثال، اندازه یا رنگ گره‌ها<sup>۷</sup> یا یال‌ها<sup>۸</sup> در شبکه را می‌توان بر اساس ویژگی‌های آن‌ها تغییر داد تا شناسایی آن‌ها در نگاه اول و همچنین درک روابط آن‌ها ساده‌تر باشد. همچنین، خوشه‌های مختلف را می‌توان با رنگ‌های متفاوت مشخص کرد. این گونه درک شبکه و اجزای آن ساده‌تر می‌شود.

- 
1. Mention
  2. Retweet
  3. Followship
  4. Centrality
  5. Betweenness
  6. Gephi
  7. Nodes
  8. Edges

هرچند استفاده از تحلیل شبکه اجتماعی درکی کلی از روابط موجود در شبکه و همچنین موقعیت برخی کاربران خاص به ما می‌دهد اما ما این روش را در تلفیق با دو روش تحلیل متنی معرفی شده بکار می‌گیریم تا علاوه بر این درک و خوانش کلی عمدتاً معطوف به کاربران، به خوانشی دقیق و جزئی از توثیت‌ها نیز برسیم.

تحلیل محتوای مردم‌نگارانه: تحلیل محتوای مردم‌نگارانه درحالی که تلاش می‌کند از نقایص روش‌های تحلیل محتوای کمی و کیفی عاری باشد، ابزاری مناسب برای تحلیل متن با توجه به بستری که متن در آن حادث می‌شود فراهم می‌کند. تحلیل محتوای کمی عمدتاً به خاطر اکتفا کردن به سنجه‌های ظاهری و ابتدایی مانند شمارش تعداد کلمات و نادیده گرفتن معنای بالقوه که در محور هم‌نشینی ممکن است ظاهر شود، موردانتقاد قرار گرفته است (کریندورف، ۲۰۰۴). تحلیل محتوای کیفی با اینکه معنای پنهان را در نیز در تحلیل وارد می‌کند، اما نمی‌تواند نقش بستری که ارتباط در آن رخ می‌دهد را شناسایی کند (شریر، ۲۰۱۴). برخلاف این دو روش، تحلیل محتوای مردم‌نگاری درحالی که از ابزارهای این دو روش استفاده می‌کند، با تلفیق آن‌ها با سنت مردم‌نگاری تلاش می‌کند بستر ارتباطی را نیز در تحلیل محتوا در نظر بگیرد. تحلیل محتوای مردم‌نگارانه، یک روش ترکیبی برای فهمیدن و مستند کردن معنای ارتباطات است (آلتهاید و اشنایدر، ۲۰۱۳). این روش در تمام مراحل خود منعطف و دایره‌ای است، به این معنا که مراحل با هم و هم‌زمان اجرا می‌شوند و بر اساس یک برنامه ساختاریافته از پیش تعیین شده پیش نمی‌رود. برخلاف تحلیل محتوای کمی که هدفش اثبات فرضیه‌ها است، هدف تحلیل محتوای مردم‌نگارانه کشف است؛ بنابراین، هیچ فرضیه‌ای را از پیش در نظر نمی‌گیرد و تلاش می‌کند تا از همه اسناد و مواد استفاده کند تا به درک درستی از محتوا برسد. این روش حرکتی دائمی و تعاملی بین مراحل مختلف در تحلیل نظیر توسعه مفاهیم، نمونه‌گیری، گردآوری داده‌ها، کدگذاری و ارائه تفسیرها است. هدف تحلیل محتوای مردم‌نگارانه در عین اینکه تلاش می‌کند نظام مند و تحلیلی بماند این است که از چارچوب‌های سفت و سخت دوری کند.

در واقع، طبقه بندی‌ها و متغیرها در اینجا نیز تحقیق را هدایت می‌کنند اما تحقیق در حیطه آن‌ها محدود نمی‌ماند. در این روش، مقولات و متغیرهای دیگر نیز می‌توانند و انتظار می‌رود- در جریان تحقیق ظاهر شوند. این روند اکتشاف دائمی و مقایسه دائمی بین موقعیت‌ها، مختصات، سبک‌ها و ظرافت‌ها ویژگی مهم تحلیل محتوای مردم‌نگارانه است. به این ترتیب، تحلیل محتوای مردم‌نگارانه برخلاف تحلیل محتوای کمی که تمرکزش عمدتاً بر اعداد و بسامدهاست، روایت‌ها را نیز در نظر می‌گیرد. همچنین تلاش این روش قرار دادن مقولات در طبقه‌های از پیش تعیین شده نیست، بلکه مقولات در هر مرحله از تحقیق می‌توانند ظاهر شوند و طرح تحقیق بر اساس آن‌ها از نو چیده شود. با استفاده از این روش، در واقع بطریق اولی روش‌های تحلیل محتوای کمی و کیفی را نیز در تحقیق دخیل کرده و از آن‌ها فراتر خواهیم رفت.

ما از تحلیل محتوای مردم‌نگارانه در تحقیق روی توئیت برای شناسایی و تحلیل ویژگی‌های مربوط به کاربران و توئیت‌ها استفاده خواهیم کرد. این روش کمک می‌کند تا بتوانیم هم ویژگی‌های کمی و عددی کاربران و توئیت‌ها را شناسایی کنیم و هم از این توصیفات کمی فراتر رفته، کردارهای شبکه‌ای آن‌ها را تحلیل کنیم. هرچند تحلیل محتوای مردم‌نگارانه بستر ارتباطی (شامل موقعیت‌ها و مختصات) را در نظر می‌گیرد اما به گفتمان‌هایی که ممکن است متن در آن‌ها تولید شده باشد یا آن‌ها را تولید کند نمی‌پردازد؛ بنابراین، هرچند این روش در سطوح مختلف برای شناسایی کردارهای شبکه‌ای سودمند است اما برای شناخت عمیق‌تر و دقیق‌تر گفتمان‌هایی که با این کردارها در ارتباطاند لازم است از روش تحلیل گفتمان نیز استفاده کنیم. ما در این تحقیق از یکی از جدیدترین رهیافت‌ها در تحلیل گفتمان انتقادی یعنی تحلیل گفتمان ا.ر.ج در تلفیق با تحلیل محتوای مردم‌نگارانه استفاده کرده تا این خلأ را رفع کنیم.

تحلیل گفتمان انتقادی رسانه‌های اجتماعی: تحلیل گفتمان ا.ر.ج یکی از رویکردهای جدید در حوزه تحلیل گفتمان است که به تحلیل گفتمان‌های تولیدشده توسط متون

منتشر شده در رسانه‌های اجتماعی و یا رابطه این متون با گفتمان‌های جامعه می‌پردازد. این رهیافت ریشه در شیوه‌های پیشین تحلیل متنی خصوصاً تحلیل گفتمان انتقادی دارد. تحلیل گفتمان انتقادی، برخلاف آنچه زبان‌شناسان اولیه می‌پنداشتند، زبان را نه به‌عنوان موجودیتی خنثی و بی‌طرف که به‌عنوان یک کنش اجتماعی در نظر می‌گیرد. تحلیل گفتمان انتقادی بر این مسئله متمرکز می‌شود که ایدئولوژی و قدرت چگونه معنا را برمی‌سازند، به زبان شکل می‌دهند و از این طریق افراد را تحت انقیاد خویش درمی‌آورند. رابطه ایدئولوژی، قدرت و زبان در این معنا دیالکتیکی، دائمی و همواره در حال تغییر است. زبان در واقع محملی برای انتقال ایدئولوژی کنشگران اجتماعی، نهادهای قدرت، رسانه‌ها و دیگر اعضای جامعه است. کنشگران با ممزوج کردن تولیدات زبانی خود با ایدئولوژی هایشان تلاش می‌کنند قدرت خویش را بر دیگران اعمال کنند؛ بنابراین تحلیلگر گفتمان انتقادی تلاش می‌کند با مطالعه نظام‌مند و دقیق زبان، این ایدئولوژی‌ها را کشف و افشا کرده و از این طریق نابرابری اجتماعی را کاهش دهد.

تعاریف و رویکردهای مختلفی برای گفتمان و تحلیل گفتمان ارائه شده است. فوکو (۱۹۸۲) گفتمان را قلمروی کلی انواع مختلف بیان می‌داند. او اضافه می‌کند که گاه با گفتمان به‌مثابه گروه‌های منفرد از انواع بیان برخورد می‌کند و گاه گفتمان را کردار نظام‌مندی می‌داند که مسئول برخی از بیان‌هاست. فرکلاف (۲۰۰۳) گفتمان را شیوه‌ای برای بازنمایی جنبه‌هایی از جهان- روندها، روابط و ساختار جهان مادی، جهان ذهنی ایده‌ها، احساسات، اعتقادات و امثال آن‌ها و جهان اجتماعی می‌داند. وود و کروگر (۲۰۰۰) در تعریفی دیگر گفتمان را به‌مثابه کردار اجتماعی‌ای می‌دانند که از تمام شکل‌های متفاوت و متنوع گفتاری و نوشتاری زبان (بخصوص کلام و متن) مشتق شده و واقعیت اجتماعی را در جامعه بازنمایی می‌کند. همچنین رویکردهای مختلفی برای تحلیل گفتمان وجود دارد که از بین آن‌ها می‌توان به رویکر فوکو، لاکلا-موفه و فرکلاف اشاره کرد (فیلیپس و

جورجنسن، ۲۰۰۲). بدلیل اینکه از شیوه‌ای دیگر برای تحلیل بهره می‌بریم و برای اجتناب از اطناب، به جزییات این رهیافت‌ها نمی‌پردازیم.

تحلیل گفتمان ا.ر.ج یک نوع تحلیل انتقادی گفتمان است که از نظر اجتماعی خود را متعهد می‌داند، مسئله‌محور و معطوف به متن است (خسروی‌نیک، ۲۰۱۷). این رویکرد برخلاف رویکردهای اصلی در تحلیل گفتمان انتقادی که از متن شروع می‌کنند و با تبیین بسترهای لایه‌های مختلف اجتماعی-سیاسی و رسانه‌ای متن ادامه می‌دهند، با مشاهده کردارها شروع می‌کند تا به متن برسد. منظور از کردار در این رویکرد، کنش‌های معین و موقعیت مندی است که افراد با استفاده از ابزارهای خاص رسانه‌ای (مانند متن نوشتاری، کامپیوتر و تلفن همراه) اجرا می‌کنند تا عضویت خود در گروه‌های اجتماعی را نمایش دهند. در اینجا، واحد تحلیل کنش رسانه‌ای شده است که همان کرداری است که متن درون آن مصرف می‌شود. این نقطه قوت و افتراق تحلیل گفتمان ا.ر.ج با تحلیل گفتمان انتقادی است.

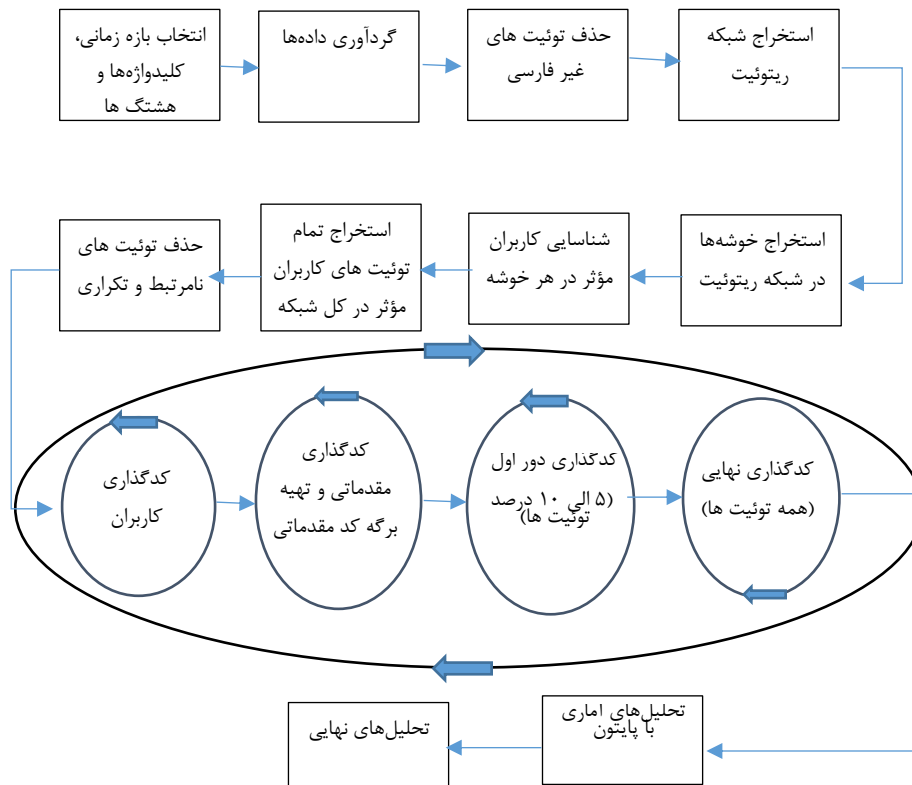
خسروی نیک (همان) معتقد است یک رهیافت مشاهده‌گرانه‌ی ارتباطی مبتنی بر کردار بیشتر با تحلیل گفتمان در رسانه‌های اجتماعی سازگار است و اطلاعات مفیدتری به دست می‌دهد. این نکته مهمی است چراکه با توجه به چرخش پارادایمی در ارتباطات رسانه‌های اجتماعی نسبت به ارتباطات جمعی، لازم است از رهیافتی استفاده کنیم که برای تحلیل کنش‌ها و کردارهای ارتباطی در این رسانه‌ها وضع شده باشد. تحلیل گفتمان ا.ر.ج این ابزار تفسیری را در اختیار ما قرار می‌دهد. همچنین این رهیافت ابزار مناسبی برای شناسایی روایت‌های شکل گرفته در توئیتر است چرا که پویایی، تکه‌تکه بودن و تنوع محتوای تولیدشده در رسانه‌های اجتماعی و به‌طور خاص توئیتر (کرمانی، ۱۳۹۸) را در نظر می‌گیرد. محتوایی که اساساً با محتوای منتشرشده در رسانه‌های جمعی متفاوت است.

خسروی نیک (۲۰۱۷) دو محور برای تحلیل گفتمان انتقادی رسانه‌های اجتماعی ترسیم می‌کند. به‌طور کلی می‌توان گفت که محور افقی به روابط درون‌متنی در دل

کردارهای متنی شکل گرفته در رسانه‌های اجتماعی می‌پردازد و محور عمودی هم خصائص خرد تحلیل متنی و هم بستر افقی را به بستر اجتماعی-سیاسی کاربران در جامعه مربوط می‌سازد. به معنی دیگر، این محور تحلیل متنی را به گفتمان‌های سیاسی-اجتماعی که در جامعه شکل گرفته‌اند وصل می‌کند. با توجه به آنچه در قسمت قبل گفتیم، محور افقی در تحلیل گفتمان ا.رج قابل ادغام با تحلیل محتوای مردمنگاره است. این دو روش در سطح افقی در کنار یکدیگر مورد استفاده قرار گرفتند تا بتوانیم اطلاعات مورد نیاز را در مورد کاربران و کردارهای ارتباطی آن‌ها که در اینجا در قالب توئیت‌هایشان بروز می‌یابد به دست آوریم. در واقع، روش تحقیق در این پژوهش مانند خود تحلیل محتوای مردم نگارانه بعنوان یکی از اجرای نظام روش شناختی این تحقیق، در همه مراحل دایره‌ای و رفت و برگشتی است؛ بنابراین برخلاف روش شناسی‌های خطی که نقطه آغاز و پایان مشخص دارند، نظام روش شناختی همانطور که ذات تحقیق روی توئیت می‌طلبد پویا، دیالکتیکی، رفت و برگشتی و در رابطه دائم با دیگر اجزای تحقیق (بستر نظری و اهداف) و اجزای خود نظام روش شناختی است.

### عملیاتی کردن رهیافت روش شناختی تحقیق در توئیت فارسی

پس از تشریح مبانی نظری رهیافت معرفی شده در این مقاله، حال به چگونگی عملیاتی کردن آن می‌پردازیم. این مراحل ابتدا در دیاگرام ۱ نشان داده شده و سپس آن‌ها را با جزئیات بیشتر شرح خواهیم داد.



دیاگرام ۱: مدل عملیاتی رهیافت روش شناختی تحقیق در توییتر فارسی

دیاگرام بالا مراحل تحقیق روی توییتر فارسی را نشان می‌دهد. این رهیافت، همان‌طور که اشاره کردیم، روش‌های محاسباتی، تحلیل‌های کیفی و متنی و روش‌های آماری کمی را به خدمت می‌گیرد تا دریافت دقیقی از توییتر فارسی (هم کاربران و هم محتوای توثیت‌ها) به دست دهد. خانه‌های مربعی در دو ردیف اول مراحل اجرای تحلیل شبکه اجتماعی را نشان می‌دهند. پس از اجرای این روش، به تحلیل‌های کیفی و متنی می‌رسیم. این مراحل هر یک، خود دایره‌ای و رفت و برگشتی هستند که به‌وسیله دایره‌ها در ردیف سوم نشان داده شده‌اند. علاوه بر این تمام این مراحل از کدگذاری کاربران گرفته تا کدگذاری نهایی توثیت‌ها نیز دایره‌ای هستند. بدین معنی که ممکن است حتی در مرحله نهایی کدگذاری توثیت‌ها، کدها و مقولات مربوط به کاربران نیز بازنگری شود. این رهیافت، نهایتاً با تحلیل



های آماری و سرانجام با ارائه تحلیل‌های نهایی تحقیق پایان می‌پذیرد. در ادامه هر یک از مراحل را با جزئیات بیشتر شرح داده و همچنین نحوه اجرایی کردن آن‌ها در انتخابات سال ۹۶ را تشریح می‌کنیم.

بازه زمانی گردآوری داده‌ها: انتخابات ریاست جمهوری سال ۱۳۹۶ در روز جمعه ۲۹ اردیبهشت برگزار می‌شد، بنابراین تصمیم گرفتیم تا گردآوری داده‌ها را از ۱۱ اردیبهشت آغاز کنیم (۱۸ روز قبل از روز انتخابات). تبلیغات رسمی کاندیداها از ۸ اردیبهشت آغاز شده بود و بنابراین بعد از گذشت ۳ روز از آغاز تبلیغات انتخاباتی، رفته‌رفته بحث‌ها درباره انتخابات بیشتر و داغ‌تر می‌شد؛ بنابراین انتخاب این تاریخ برای شروع گردآوری توئیت‌ها منطقی بود. گردآوری توئیت‌ها بعد از برگزاری انتخابات تا ۴ خرداد ادامه یافت. بعد از گذشت چندین روز از انتخابات و اعلام رسمی نتایج در ۳۰ اردیبهشت، داده‌ها کفایت لازم برای تحلیل کردارهای شبکه‌ای کاربران را داشتند هرچند روشی علمی برای اطمینان کامل از کفایت چنین داده‌هایی وجود ندارد، تلاش کردیم با انتخاب بازه زمانی وسیع (۲۵ روز) که پیش از روز انتخابات شروع شده و تا پس از آن ادامه داشته است اطمینان نسبی از این امر به دست آوریم؛ بنابراین گردآوری توئیت‌ها متوقف شد.

انتخاب کلیدواژه‌ها و هشتگ‌ها: برای گردآوری توئیت‌ها، ابتدا فهرستی شامل نام همه کاندیداها (با هشتگ و به صورت کلیدواژه با نگارش‌های متفاوت آن‌ها) تهیه کردیم و در ادامه به صورت گلوله برفی (بی، ۱۳۸۴)، با ظاهر شدن هشتگ‌ها و کلیدواژه‌های جدید، آن‌ها را به فهرست مورد نظر اضافه کردیم تا توئیت‌هایی که از این هشتگ‌ها و کلیدواژه‌ها نیز استفاده کرده‌اند گردآوری شود. در نهایت، ۹۴ هشتگ و کلیدواژه برای گردآوری توئیت‌ها استفاده شد. فهرست این کلیدواژه‌ها و هشتگ‌ها در ضمیمه ۱ ارائه شده است. البته در مورد این فهرست چند نکته باید مدنظر قرار گیرد.

اول اینکه در مواردی که تنها کلیدواژه -بدون هشتگ- در فهرست ذکر شده است نیز، هم توئیت‌هایی که حاوی این کلیدواژه‌ها بوده‌اند و هم توئیت‌هایی که حاوی این

کلیدواژه‌ها به همراه هشتگ بوده‌اند توسط DMI-TCAT گردآوری شدند. در مواردی که کلیدواژه‌های مدنظر جزو افعال یا کلمات رایج در زبان فارسی بودند (مانند تغییر) از هشتگ در ابتدای آن‌ها استفاده شد تا توئیت‌های غیر مرتبط گردآوری نشود. مورد دوم اینکه، به دلیل اینکه برخی کاربران با استفاده از دو هشتگ انگلیسی #IranElection و #IranElections2017 به زبان فارسی توئیت می‌کردند این هشتگ‌ها را در فهرست قراردادیم اما توئیت‌هایی که به زبان انگلیسی بودند در مرحله پایش داده‌ها حذف شدند. گردآوری داده‌ها: برای گردآوری داده‌های تحقیق از API<sup>۱</sup> توئیت استفاده کردیم. توئیت در واقع به برنامه نویسان و دیگر افراد این اجازه را می‌دهد که با استفاده از API های مختلف این شبکه، داده‌های موردنیاز خود را از توئیت دریافت کنند. البته دریافت توئیت‌ها به‌طور رایگان تنها به ۷ روز محدود است؛ یعنی هنگامی که درخواستی به API توئیت ارسال می‌شود، این درخواست تنها می‌تواند توئیت‌های منتهی به هفت روز پیش از آن را دریافت کند. دریافت توئیت‌های تاریخی و توئیت‌هایی که قبل از آن ۷ روز ارسال شده‌اند مستلزم پرداخت هزینه به توئیت است.

برای دریافت توئیت‌ها و ارسال درخواست به API توئیت از ابزارهای مختلفی می‌توان استفاده کرد. این درخواست‌ها به‌طور مستقیم و با استفاده از زبان‌های برنامه‌نویسی می‌تواند ارسال شود. همچنین می‌توان از ابزارهای دیگر که نیاز به مهارت‌های برنامه‌نویسی ندارند استفاده کرد. چنین ابزارهایی هم به‌صورت رایگان و هم به‌صورت پولی موجود هستند که از آن‌ها می‌توان به DiscoverText، Netlytic و NodeXL اشاره کرد. نسخه ۱۱ نرم‌افزار Nvivo نیز امکان دریافت توئیت را فراهم کرده است. ما برای دریافت توئیت‌ها از زبان برنامه‌نویسی آر و DMI-TCAT<sup>۲</sup> استفاده کردیم. DMI-TCAT ابزاری است که به افراد این امکان را می‌دهد تا داده‌های معتبر و قابل تکثیر از توئیت دریافت کنند و استفاده از آن

---

1. Application programming interface

2. Digital Methods Initiative Twitter Capture and Analysis Toolset

برای دریافت داده‌ها از توئیتر در علوم اجتماعی نیز مرسوم است (بورا و رایدر، ۲۰۱۴). برای این تحقیق، در مدت‌زمان ۲۵ روز و بر اساس ۹۴ کلیدواژه و هشتگ، دو میلیون و ۵۹۶ هزار و ۲۸۴ توئیت که شامل یک یا چند مورد از این هشتگ‌ها و کلیدواژه‌ها بود گردآوری شد.

پس از اتمام گردآوری توئیت‌ها، با دریافت خروجی از DMI-TCAT به صورت فایل TSV توئیت‌ها آماده پالایش بودند. لازم به ذکر است که خروجی DMI-TCAT می‌تواند در فرمت‌های مختلف صادر شود. دو نوع از بهترین فرمت‌ها برای فایل توئیت‌ها CSV و TSV است اما فرمت CSV به دلیل اینکه فیلدهای مختلف را بر اساس ویرگول (،) از هم جدا می‌کند می‌تواند مشکل‌ساز باشد. به عنوان مثال هنگامی که یک برنامه مانند اکسل می‌خواهد فایل داده‌ها را بخواند، ستون‌ها را بر اساس ویرگول جدا می‌کند؛ بنابراین ستون مربوط به توئیت و نام کاربر با توجه به ویرگولی که بین آن‌ها در فایل CSV قرار دارد جدا می‌شوند اما در این حالت اگر خود متن توئیت هم ویرگول داشته باشد، توئیت در دو ستون تفکیک خواهد شد و این‌گونه فایل نهایی دچار مشکل می‌شود؛ بنابراین بهتر است از فرمت TSV استفاده کنیم که ستون‌ها را بر اساس برگ از هم جدا می‌کند.

پالایش داده‌ها: پالایش داده‌ها در چند مرحله انجام شد. ابتدا باید توئیت‌هایی که به زبانی غیر از زبان فارسی نوشته شده بود حذف می‌شدند چرا که تمرکز این تحقیق بر توئیت‌های فارسی بود. در مرحله بعد باید شبکه ریتوئیت استخراج می‌شد و در نهایت فیلدهای غیر ضروری حذف می‌شدند تا تنها فیلدهایی که در تحقیق کاربرد دارند در فایل مربوطه باقی بمانند. برای انجام این مراحل، از نرم‌افزار Tableau و همچنین OpenRefine استفاده شد. هر دو نرم‌افزار قابلیت کار با داده‌های کلان را دارند و به نحو مطلوب می‌توانند برای پالایش داده‌ها استفاده شوند.

با حذف توئیت‌هایی که به زبان‌های دیگر نوشته شده بود، مشخص شد که تعداد توئیت‌ها به زبان فارسی ۲۱۲۱۶۴۲ توئیت است. سپس دو شبکه ریتوئیت و منشن با توجه به

شبکه توئیت‌های فارسی استخراج شد. جدول ۱ تعداد توئیت‌های هر شبکه را نشان می‌دهد.

جدول ۱: مشخصات شبکه‌های ریتوئیت و منشن

تعداد توئیت‌ها	نام شبکه
۱۲۰۸۷۲۳	ریتوئیت
۱۱۱۷۱۰	منشن

جدول ۱ نشان می‌دهد که تعداد توئیت‌های شبکه ریتوئیت یک میلیون و ۲۰۸ هزار و ۷۲۳ توئیت و تعداد توئیت‌های شبکه منشن صد و یازده هزار و هفتصد و ده توئیت بوده است. همچنین، هشتصد و یک هزار و دویست و نه توئیت یک<sup>۱</sup> در شبکه وجود داشت. تعداد گره‌ها در شبکه ریتوئیت ۶۲۶۳۳ و تعداد یال‌ها نیز ۷۱۳۶۹۶ بود. در اینجا وجود یال از کاربر آ به کاربر ب به این معنی بود که کاربر آ توئیتی از کاربر ب را ریتوئیت کرده است. در مرحله بعد، فیله‌های غیرضروری حذف شدند تا تنها فیله‌های موردنیاز در فایل نهایی باقی بمانند. این فیله‌ها عبارت بودند از: نام کاربری، توضیحات پروفایل کاربر، متن توئیت و تعداد فالوئرهای کاربر. علاوه بر این فیله‌ها، برخی فیله‌های دیگر نیز برای تحلیل موردنیاز بود که آن‌ها را در مرحله تحلیل کیفی معرفی کرده‌ایم. در قسمت مربوط به توضیح چگونگی انجام تحلیل‌های کیفی توضیحات لازم در این مورد ارائه خواهد شد.

تحلیل‌های مربوط به شبکه شامل دو مرحله بود. در مرحله اول خوشه‌های شکل گرفته در شبکه ریتوئیت را با استفاده از الگوریتم تعیین خوشه‌ها در شبکه‌های بزرگ بر اساس شاخص ماژولاریتی<sup>۲</sup> شناسایی کردیم. سپس با استفاده از سنجه PageRank، مؤثرترین کاربران در هر خوشه را مشخص کردیم. جزییات مربوط به هر مرحله در ادامه ارائه شده است.

1. Original  
2. Modularity

استخراج خوشه‌ها در شبکه ریتوئیت: ماژولاریتی یک سنجه مناسب برای تعیین جوامع<sup>۱</sup> در شبکه‌های بزرگ است که مشخص می‌کند آن شبکه شامل چند خوشه است و همچنین هر یک از این خوشه‌ها را شناسایی می‌کند (بلاندل و همکاران، ۲۰۰۸). برای تعیین جوامع در شبکه، مفاهیم مختلفی وجود دارد. از جمله این مفاهیم می‌توان به اجتماع<sup>۲</sup>، قطعه<sup>۳</sup>، محفل<sup>۴</sup>، دایره<sup>۵</sup> و خوشه<sup>۶</sup> اشاره کرد (اسکات، ۲۰۰۰). قطعه، محفل و دایره عمدتاً برای شناسایی گروه‌های کوچک با مراتب مختلف بکار می‌رود (اسکات، ۲۰۰۰) که در این تحقیق مدنظر نیست. همچنین اجتماع و خوشه نیز از نظر مفهومی بسیار به یکدیگر نزدیک هستند و در مواردی بجای هم بکار می‌روند. خوشه‌ها در معنای کلی بخش‌هایی هستند که به‌طور طبیعی درون شبکه پدیدار شده‌اند (کادوشین، ۲۰۰۲). خوشه‌ها در مجموع باید دو شرط ضروری داشته باشند. اول اینکه گره‌های درون یک خوشه باید دارای ارتباطات قوی با یکدیگر بوده و دوم، گره‌های یک خوشه ارتباطات اندکی با گره‌های خوشه‌های دیگر داشته باشند. در این معنا، خوشه‌ها بر اساس مفهوم هوموفیلی شکل می‌گیرند، یعنی کاربرانی که نزدیکی بیشتر با یکدیگر از جنبه‌های مختلف دارند به‌طور طبیعی در شبکه با یکدیگر روابط قوی‌تری خواهند داشت و یک خوشه را تشکیل می‌دهند (هیملبوین، ۲۰۱۴). شاخص ماژولاریتی مشخص می‌کند که چگونه گره‌ها در شبکه می‌توانند به‌طور طبیعی در بخش‌های مختلف قرار بگیرند و این بخش‌ها - که ما آن‌ها را خوشه می‌نامیم - را مشخص می‌کند.

در این تحقیق، شاخص ماژولاریتی توسط نرم‌افزار گفی محاسبه شده و خوشه‌ها نیز توسط این نرم‌افزار مشخص شدند. در شبکه ریتوئیت، شاخص ماژولاریتی برابر با ۰,۵۵ بود. لازم به ذکر است که شاخص ماژولاریتی همواره بین ۰-۱ قرار دارد و هر چه این

- 
1. Community detection
  2. Community
  3. Component
  4. Clique
  5. Circle
  6. Cluster

شاخص بزرگ‌تر باشد به این معناست که مرز بین خوشه‌ها واضح‌تر است اما اگر این شاخص کمتر از ۰,۵ باشد، آنگاه تعیین مرز بین خوشه‌ها دشوار می‌شود. نرم‌افزار گفنی ۱۰۰۶ خوشه در شبکه ریتوئیت بر اساس ماژولاریتی شناسایی کرد. از این خوشه‌ها تنها ۵ خوشه بود که بیشتر از ۵ درصد از حجم کل شبکه را شامل می‌شدند و اندازه خوشه‌های دیگر بسیار کوچک بود؛ بنابراین تنها این ۵ خوشه را در نظر گرفته و از خوشه‌های دیگر صرف‌نظر کردیم. تحلیل اولیه و مطالعه توئیت‌های هر خوشه به‌طور تصادفی نشان داد که بزرگ‌ترین خوشه مربوط به اصلاح‌طلبان، دومین خوشه متعلق به اصولگرایان و خوشه سوم از نظر بزرگی مربوط به ایرانیان مهاجر (مهاجران) است. خوشه بعدی متعلق به سازمان مجاهدین خلق بود و خوشه آخر نیز عمدتاً از بات‌ها تشکیل شده بود. علاوه بر این خوشه، تحقیقات قبلی پارسی و دیگران، (۲۰۱۸) و تحلیل‌های مقدماتی نشان می‌دهد که خوشه مربوط به مجاهدین خلق نیز عمدتاً از بوت‌ها تشکیل شده و آن‌ها تلاش می‌کنند با استفاده از این بوت‌ها هشتگ‌های مورد نظر خود را ترند کرده و جریان محتوا را در توئیت فارسی در دست بگیرند. هر چند عملکرد این بوت‌ها می‌تواند موضوع تحقیقاتی مستقل باشد اما مطالعه آن‌ها هدف این تحقیق نبود، بنابراین این دو خوشه را از تحلیل‌ها کنار گذاشته و به تحلیل سه خوشه اول اکتفا کردیم. جدول ۲ اطلاعات مربوط به این خوشه‌ها را نشان می‌دهد.

جدول ۲: خوشه‌ها در شبکه ریتوئیت

نام خوشه	درصد از کل شبکه	تعداد گروه‌ها	تعداد یال‌ها
اصلاح‌طلبان	۵۲,۹۳	۳۳۱۵۷	۴۵۱۴۹۹
اصولگرایان	۱۲,۵	۷۸۲۷	۱۱۱۲۴۸
ایرانیان مهاجر	۷,۴	۳۶۶۵	۱۷۵۸۲

شناسایی کاربران مؤثر در خوشه‌ها: سنجه‌های مختلفی برای شناسایی کاربران مؤثر در شبکه وجود دارد که سنجه‌های مرکزیت<sup>۱</sup> نامیده می‌شوند. در واقع، این سنجه‌ها محبوبیت<sup>۲</sup>، فعالیت<sup>۳</sup> و پدیداری<sup>۴</sup> کاربران را می‌سنجند. در گذشته به سنجه‌های ساده‌ای مانند تعداد فالوئرهای یک کاربر برای سنجش محبوبیت یا تعداد توئیت‌ها برای سنجش فعالیت اکتفا می‌شد اما این سنجه‌ها، با توجه به این حقیقت که تعداد بالاتر فالوئر لزوماً نشان‌دهنده اثرگذاری بیشتر نیست و یا با رواج بوت‌ها و حساب‌های کاربری تقلبی، این سنجه‌ها دیگر نمی‌توانند مبنای قابل‌اعتمادی برای سنجش و شناسایی میزان محبوبیت، فعالیت و پدیداری کاربران باشند (برانس و استیگلitz، ۲۰۱۳). با توجه به این مسئله، سنجه‌های جدیدی برای تعیین این موارد پیشنهاد شده است. از جمله این سنجه‌ها می‌توان به شاخص اندازه<sup>۵</sup>، بینیت<sup>۶</sup>، نزدیکی<sup>۷</sup>، eigenvector و PageRank اشاره کرد.

مبنای تعیین شاخص اندازه بر اساس تعداد یال‌هایی است که به یک گره وارد شده (اندازه ورودی<sup>۸</sup>) یا از آن خارج می‌شود (اندازه خروجی<sup>۹</sup>). این دو اندازه در شبکه‌های بی‌جهت<sup>۱۰</sup> یکسان هستند اما در شبکه‌های جهت‌دار<sup>۱۱</sup> با یکدیگر تفاوت دارند. در حالت عادی در توئیت، شاخص اندازه خروجی برای سنجش میزان فعالیت کاربر استفاده می‌شود چراکه این شاخص به این معناست که کاربر چند توئیت ارسال کرده است (برانس و استیگلitz، ۲۰۱۳). در شبکه ریتوئیت، اندازه خروجی نشان‌دهنده تعداد توئیت‌هایی است که کاربر ریتوئیت کرده است و اندازه ورودی نیز نشان می‌دهد که توئیت‌های کاربر چند

- 
1. Centrality
  2. Popularity
  3. Activity
  4. Visibility
  5. Degree
  6. Betweenness
  7. Closeness
  8. In degree
  9. Out degree
  10. Undirected
  11. Directed

بار ریتوئیت شده‌اند؛ بنابراین در شبکه ریتوئیت، اندازه خروجی نشان‌دهنده میزان فعالیت و اندازه ورودی نشان‌دهنده میزان پدیداری<sup>۱</sup> کاربر است. با این همه این شاخص با همه لینک‌های ورودی و خروجی یکسان برخورد می‌کند و کیفیت‌های متفاوت آن‌ها را در نظر نمی‌گیرد.

پیش از تعریف شاخص بینیت و نزدیکی لازم است ابتدا مفهوم مسیر را توضیح دهیم چراکه این دو شاخص بر اساس مسیرهایی که از آن‌ها می‌گذرد محاسبه می‌شوند. مسیر در یک گراف که بازنمایی هندسی یک شبکه است، کوتاه‌ترین راه ممکن برای رسیدن از گره آ به گره ب است (اسکات، ۲۰۰۰). در واقع، راه‌های مختلفی برای رسیدن به گره ب از گره آ وجود دارد، کوتاه‌ترین آن‌ها مسیر نامیده می‌شود. شاخص بینیت، تعداد مسیرهایی که از گره آ می‌گذرد را محاسبه می‌کند؛ بنابراین هر چه بینیت یک گره بیشتر باشد، آن گره بین مسیرهای بیشتری قرار دارد. شاخص بینیت برای گره آ از تقسیم تعداد مسیرهایی که از آن گره می‌گذرد بر تعداد کل مسیرهای درون شبکه به دست می‌آید (همان). هر چه شاخص بینیت یک گره بالاتر باشد، حذف آن باعث می‌شود شبکه بیشتر دچار فروپاشی شود.

شاخص نزدیکی به بیان ساده به این معناست که یک گره از گره‌های دیگر در کل شبکه چقدر فاصله دارد (کادوشین، ۲۰۰۲). این شاخص در واقع میانگین طول تمام مسیرهای ممکن از گره آ به تمام گره‌های درون شبکه است. این شاخص نشان می‌دهد که برای رسیدن از یک گره به گره‌های دیگر در شبکه چه فاصله‌ای باید طی شود. به طور منطقی هر چه این فاصله کمتر باشد، بیانگر این است که آن گره میزان دسترسی<sup>۲</sup> بالاتری به کل شبکه دارد، بنابراین انتشار اطلاعات سریع‌تر توسط آن گره انجام می‌شود.

شاخص eigenvector توسط پیتر گولد معرفی شده و به شاخص گولد هم معروف است (گولد، ۱۹۶۷). این شاخص بر اساس مرکزیت اندازه تعریف می‌شود و بر این فرض

---

1. Visibility

2. Reach



بنا شده که مرکزیت هر گره وابسته به مجموع مرکزیت گره‌هایی است که با آن مرتبط هستند. در واقع، یک گره وقتی مهم است که با گره‌های مهم در رابطه باشد. بدین ترتیب ممکن است تعداد ارتباطات یک گره کم باشد (شاخص اندازه پایین) اما آن گره با گره‌های مهم شبکه در رابطه باشد و در نتیجه شاخص eigenvector آن بالا باشد.

هرچند شاخص eigenvector، تعداد روابط را در نظر نمی‌گیرد و تنها به کیفیت گره‌های مرتبط کار دارد، این مشکل در شاخص PageRank حل شده است. شاخص PageRank در واقع هم کیفیت گره‌های مرتبط و هم تعداد روابط را در نظر می‌گیرد تا از این طریق بتواند مهم‌ترین گره‌ها در یک شبکه را شناسایی کند؛ بنابراین استفاده از شاخص PageRank به‌طور منطقی بر استفاده از eigenvector ارجحیت دارد. شاخص PageRank که ابتدا برای ارزیابی وبسایت‌های اینترنتی وضع شده بود در معنای ساده مانند مایعی است که درون شبکه از طریق یال‌ها جریان می‌یابد و گره‌های مهم‌تر را بیشتر پر می‌کند (ایشلی و کلینبرگ، ۲۰۱۰). در این تحقیق با استفاده از شاخص PageRank مؤثرترین کاربران در هر خوشه (۵۰ کاربر در هر خوشه) شناسایی شدند تا با استفاده از توئیت‌های آن‌ها، تحلیل‌های کیفی انجام شود.

استخراج توئیت‌ها: پس از شناسایی کاربران مؤثر در شبکه ریتوئیت (به‌عنوان شبکه انتشار اطلاعات) بر اساس PageRank، تمام توئیت‌های این کاربران در کل شبکه (شامل توئیت‌های اصلی، ریتوئیت و منشن‌های این کاربران) استخراج شد. استخراج توئیت‌ها توسط نرم‌افزار Tableau انجام شد. با توجه به اینکه تحقیقات پیشین که تشابه زیادی با این تحقیق دارند عمدتاً برای انجام تحلیل‌های کیفی بر بیست کاربر مؤثر تمرکز کرده‌اند (پاپاچاریسی، ۲۰۱۵)، ما تعداد ۵۰ کاربر مؤثر در هر شبکه را در نظر گرفتیم تا داده‌های بیشتری برای تحلیل گردآوری شود. جدول ۳ تعداد توئیت‌های این کاربران را نشان می‌دهد.

جدول ۳: تعداد توئیت‌های ۵۰ کاربر مؤثر در هر خوشه

نام خوشه	تعداد توئیت‌های ۵۰ کاربر مؤثر
اصلاح‌طلبان	۱۴۸۵۶
اصولگرایان	۱۰۶۲۸
ایرانیان مهاجر	۵۶۱۴
مجموع	۳۱۰۹۸

مجموع توئیت‌های این ۱۵۰ کاربر در کل شبکه (حجم جامعه آماری) ۳۱۰۹۸ توئیت بود. با توجه به محدودیت زمان و هزینه در چنین تحقیقاتی و همچنین وقت‌گیر بودن کدگذاری کیفی برای این تعداد توئیت، برای انتخاب نمونه نهایی توئیت‌ها از روش نمونه‌گیری تصادفی خوشه‌ای استفاده کردیم. با استفاده از این روش، حجم نمونه نهایی متناسب با حجم هر همگان خواهد بود. با استفاده از نرم‌افزار PASS نسخه ۱۵ حجم نمونه نهایی در سطح اطمینان ۹۹ درصد با حاشیه خطای ۱ درصد، ۱۰۸۴۱ توئیت بود. در واقع سعی کردیم با انتخاب حداقل حاشیه خطا و بیشترین سطح اطمینان، بالاترین میزان اطمینان را از نتایج داشته باشیم و تا حد ممکن اعتبار تحقیق را افزایش دهیم. نمونه نهایی به‌طور احتمالی و تصادفی بر اساس حجم هر همگان انتخاب شد.

در اینجا باید به یک نکته دیگر اشاره کنیم و آن‌هم وجود توئیت‌های نامرتبط و نامشخص در داده‌ها بود. با وجود اینکه توئیت‌ها از API توئیت‌ها به‌طور مستقیم گردآوری شده بودند اما برخی از توئیت‌ها ناقص بوده و متن کامل آن‌ها موجود نبود که می‌تواند به دلیل مشکلات فنی در نرم‌افزار TACT یا API توئیت‌ها رخ داده باشد. در این موارد، با مراجعه به توئیت و جستجوی توئیت تلاش می‌کردیم توئیت کامل را پیدا کنیم اما در صورتی که توئیت حذف شده بود یا پیدا نمی‌شد، آن را با عبارت نامشخص کدگذاری می‌کردیم. همچنین از آنجا که انتخابات شوراهای شهر و روستا نیز هم‌زمان با انتخابات ریاست جمهوری برگزار می‌شد، برخی توئیت‌های مربوط به این انتخابات نیز در داده‌ها وجود داشت که این توئیت‌ها با کد نامرتبط دسته‌بندی و از تحلیل‌ها کنار گذاشته شدند. لازم به

ذکر است که حذف توئیت‌های نامرتب و ناقص می‌تواند پس از پالایش داده‌ها صورت گیرد ولی با توجه به اینکه نرم‌افزارها و زبان پایتون با زبان فارسی به‌طور کامل هماهنگ نیستند احتمال خطا در استفاده از روش‌های خودکار برای این کار وجود دارد، به همین دلیل ما در مرحله کدگذاری توئیت‌های ناکارآمد را حذف می‌کنیم تا اطمینان بالاتری در مورد نتایج وجود داشته باشد. در مجموع ۴۲۵ توئیت نامشخص و نامرتب از تحلیل‌ها کنار گذاشته شد. جدول ۴ سهم هر همگان در نمونه نهایی را پیش و پس از حذف این توئیت‌ها نشان می‌دهد.

جدول ۴: حجم نمونه نهایی و سهم هر همگان

نام خوشه	تعداد توئیت‌های نمونه پیش از حذف توئیت‌های نامشخص و نامرتب	تعداد توئیت‌های نمونه پس از حذف توئیت‌های نامشخص و نامرتب
اصلاح‌طلبان	۵۵۶۷	۵۳۷۲
اصولگرایان	۳۴۶۲	۳۳۴۳
ایرانیان مهاجر	۱۸۱۲	۱۷۰۱
مجموع	۱۰۸۴۱	۱۰۴۱۶

بنابراین حجم نهایی نمونه که برای تحلیل‌های کیفی آماده شد، ۱۰۴۱۶ توئیت بود. با مشخص شدن و انتخاب توئیت‌ها برای تحلیل‌های کیفی، مرحله تحلیل شبکه اجتماعی به پایان رسیده وارد تحلیل‌های متنی می‌شویم. تحلیل‌های متنی و کیفی در این رهیافت دو قسمت اساسی دارد: کدگذاری کاربران و کدگذاری توئیت‌ها. در ادامه جزئیات مربوط به هر مرحله را تشریح می‌کنیم ولی پیش از آن لازم است شیوه کدگذاری تشریح شود.

شیوه کدگذاری: برای کدگذاری در هر قسمت (کاربران و توئیت‌ها)، ابتدا در دو مرحله فهرست نهایی کدها مشخص شده و در مرحله نهایی تمام کاربران و توئیت‌ها (به‌طور جداگانه) بر اساس این فهرست کدگذاری شدند. مرحله اول کدگذاری بر اساس کدگذاری موقتی<sup>۱</sup> پیش رفت که سال‌دانا آن را جزو روش‌های اکتشافی دسته‌بندی کرده

---

1. Provisional

است (سالدانا، ۲۰۱۵). در مرحله دوم نیز از کدگذاری الگویی<sup>۱</sup> که سالدانا (همان) برای کدگذاری مرحله دوم معرفی کرده استفاده شد. هر دوی این روش‌ها به دلیل پویایی‌ای که در خود دارند و در هر مرحله می‌توانند دستخوش تغییر و بازنگری شوند با رهیافت روش‌شناختی تحقیق مبتنی بر تحلیل محتوای مردم‌نگارانه و تحلیل گفتمان ا.رج همخوانی دارند. جزئیات کدگذاری در هر دو مرحله در ادامه ارائه می‌شود.

برای کدگذاری در این روش از ۳ کدگذار استفاده شد. کدگذاران دانشجوی مقطع دکترا و کارشناسی ارشد رشته علوم ارتباطات اجتماعی و با مبانی نظری کار آشنا بودند. علاوه بر این، در ۴ جلسه روند کدگذاری و مبانی کار به‌طور کامل برای آن‌ها تشریح شد و در جریان کدگذاری نیز در یک روند دائمی اعضای تیم کدگذاری از طریق نرم‌افزار Slack سؤال‌ها و ابهامات خود را مطرح می‌کردند و بحثی ثابت درمورد کدگذاری در جریان بود تا همگرایی بیشتری بین کدگذارها حاصل شود. تحلیل توئیت‌های فارسی و کدگذاری آن‌ها با توجه به جنس توئیت‌ها در توئیت‌های فارسی نسبت به کدگذاری متون رسمی مانند مطالب رسانه‌ای پیچیدگی بیشتری دارد. به همین دلیل در این رهیافت برخلاف اغلب رهیافت‌ها به دو کدگذار اکتفا نشده و تعداد کدگذار بیشتری بکار گرفته می‌شود. همچنین در این رهیافت محقق جدا از کدگذاران بر کار نظارت کرده و یک نفر نیز بین محقق و کدگذاران مرحله مقوله‌بندی را انجام می‌دهد؛ بنابراین به این صورت اعتبار کدگذاری بالاتر می‌رود.

در کدگذاری موقتی، یک فهرست ابتدایی<sup>۲</sup> از کدها بر اساس تجربیات، دانش و اطلاعات زمینه‌ای محقق و کدگذاران و همچنین بستر نظری تهیه می‌شود. با مشورت بین کدگذاران این فهرست ابتدایی تهیه شد که در مرحله اول کدگذاری مورداستفاده قرار گرفت. فهرست ابتدایی در روند کدگذاری مرحله اول به موازات کدگذاری و تحلیل داده‌ها، بازنگری می‌شود که این می‌تواند شامل اصلاح، حذف و یا توسعه فهرست برای

---

1. Pattern  
2. Start list

پذیرفتن کدهای جدید باشد (همان، ص ۱۴۴). در این مرحله، کدگذاران با مطالعه تصادفی توئیت‌ها و برگه کد اولیه پیشنهادهای خود را مطرح کرده تا به برگه کد ابتدایی اصلاح‌شده مورد توافق تمام کدگذاران برسیم. در مرحله دوم از کدگذاری الگویی استفاده شد چراکه این روش برای شناخت قالب‌ها و همچنین روایت‌ها به‌عنوان الگوهای شکل‌گرفته در توئیت‌ها مناسب‌تر است. در این مرحله، به هر یک از کدگذاران ۵۰۰ توئیت (در مجموع ۱۵۰۰ توئیت) اختصاص داده شد تا با کدگذاری آن‌ها بتوانیم به فهرست نهایی برای کدگذاری دست پیدا کنیم. در اینجا لازم است توضیح دهیم که تمام توئیت‌ها ابتدا به‌طور کاملاً تصادفی مرتب شدند تا هیچ سوگیری‌ای در اختصاص توئیت‌ها وجود نداشته باشد؛ بنابراین، کدگذاران توئیت‌هایی از هر سه خوشه را کدگذاری می‌کردند. پس از مشخص شدن فهرست نهایی کدها، تمام کاربران و توئیت‌ها بر اساس این فهرست کدگذاری شدند. البته در تمامی این مراحل کدگذاران آزاد بودند تا در صورت پدیدار شدن الگوهای جدید یا ناکارآمدی الگوهای قدیمی برگه را اصلاح کنند. در مورد انتخاب توئیت‌های برای کدگذاری در این مرحله نیز پیشنهاد ما انتخاب حداقل ۵ الی ۱۰ درصد توئیت‌ها به‌طور تصادفی است. این کار باعث می‌شود گستره معقولی از توئیت‌ها کدگذاری شده و برگه کد تهیه‌شده فاصله زیادی با برگه نهایی نداشته باشد. انتخاب دقیق تعداد توئیت‌ها، به اهداف تحقیق و همچنین مهارت کدگذاران بستگی دارد.

کدگذاری کاربران: برنامه عمل برای کدگذاری کاربران شامل چند مرحله بود. در مرحله اول کدگذاران تلاش می‌کردند کاربر را در توئیت‌شناسایی کنند. به این منظور، ابتدا نام کاربری کاربر در توئیت جستجو می‌شد تا با مراجعه به نمایه کاربر کدگذاری انجام شود. در این مرحله ممکن بود نام کاربری در جستجو پیدا نشود که به این معنی بود که یا کاربر حساب کاربری خود را پاک کرده یا نام کاربری‌اش را تغییر داده است. به همین دلیل در مرحله بعد، متن توئیت در توئیت‌جستجو می‌شد. در این حالت اگر توئیت پیدا می‌شد می‌توانستیم به کاربر دست پیدا کنیم و در برگه کد نیز مشخص می‌کردیم که کاربر نام

کاربری خود را تغییر داده است. اگر با جستجوی متن کامل به توثیق نمی‌رسیدیم، بخشی از توثیق را جستجو می‌کردیم چراکه در برخی موارد نرم‌افزارها در تشخیص بعضی حروف فارسی و عربی دچار اشتباه می‌شوند (مانند «ی»)، به همین دلیل تلاش می‌کردیم تا با جستجوی قسمتی از توثیق که این حروف را نداشته باشد به توثیق اصلی و از آن طریق به کاربر برسیم؛ بنابراین مراحل شناسایی کاربر به شرح زیر بود:

۱- جستجوی نام کاربر در تویتر

۲- جستجوی متن توثیق در تویتر

۳- جستجوی بخشی از متن توثیق در تویتر

اگر هیچ‌یک از این مراحل به نتیجه نمی‌رسید، نتیجه‌گیری منطقی این بود که کاربر حساب کاربری خود را حذف کرده است. در صورت حذف حساب کاربری کاربر، این مورد در برگه کدگذاری قید می‌شد و تلاش می‌کردیم براساس داده‌های دریافت شده از DM-TCAT تا حد ممکن کار کدگذاری را انجام دهیم.

در صورتی که حساب کاربری کاربر وجود داشت، ابتدا تلاش می‌کردیم بر اساس داده‌های دریافت شده و همچنین اطلاعات موجود در نمایه کاربر کدهای صحیح را به کاربر اختصاص دهیم. در برخی موارد که کاربر جزو چهره‌های شناخته‌شده بود در همان مرحله اول می‌شد بسیاری از اطلاعات موردنیاز را مشخص کرد اما در صورتی که کاربر چهره شناخته‌شده‌ای نبود، از اطلاعات نمایه او مانند شرح حال، عکس و نام کاربری برای تکمیل کدها استفاده می‌کردیم. در صورتی که این اطلاعات کافی نبود، با مطالعه توثیق‌های کاربر تلاش می‌کردیم کدهای صحیح را انتخاب کنیم. در صورتی که مطالعه توثیق‌های کاربر نیز به ما اطلاعات کافی نمی‌داد، به شبکه روابط او توجه می‌کردیم و تلاش می‌شد با توجه به افرادی که کاربر با آنها در ارتباط است کدهای درست را انتخاب کنیم. در نهایت، اگر هیچ‌یک از این روش‌ها به نتیجه نمی‌رسید تلاش می‌شد حساب‌های کاربر در رسانه‌های اجتماعی دیگر خصوصاً تلگرام پیداشده و بر اساس آنها بتوانیم کار

کد گذاری را انجام دهیم. تمامی این مراحل منطبق با روش‌های تحلیل محتوای مردم‌نگارانه و تحلیل گفتمان ا.ر.ج بود؛ بنابراین مراحل کد گذاری کاربران به شرح زیر بود با این شرح که اگر هر یک از مراحل به نتیجه نمی‌رسید به مرحله بعد می‌رفتیم:

۱- کاربر چهره مشهوری است؟

۲- اطلاعات نمایه کاربر

۳- خواندن توئیت‌های کاربر

۴- شناسایی و تحلیل شبکه روابط کاربر

۵- پیدا کردن حساب‌های کاربر در رسانه‌های اجتماعی دیگر

در نهایت برای کاربران توئیت فارسی، چند طبقه و برای هر طبقه مقولات مرتبط شناسایی شد که می‌تواند در تحقیقات بعدی مورد استفاده قرار بگیرد. این موارد در ضمیمه شماره ۱ ارائه شده‌اند.

کد گذاری توئیت‌ها: پس از کد گذاری کاربران، در مرحله بعدی بایستی توئیت‌های نمونه تحقیق کد گذاری شود تا آماده تحلیل‌های آماری شویم. ما بر اساس اهداف تحقیق برخی متغیرها را تعریف کردیم که باید کد گذاری می‌شدند. لازم به ذکر است که این متغیرها مانند مقولات مربوطه در هر مرحله می‌توانستند اصلاح و بازنگری گردند. تصویر زیر شمایی کلی از متغیرها و کد گذاری توئیت‌ها را نشان می‌دهد. البته باید توجه کرد که بسته به اهداف تحقیق، این متغیرها می‌توانند دستخوش تغییر شوند و برای تحقیق در مورد کنشگری سیاسی در توئیت ما این متغیرها را پیشنهاد می‌کنیم.

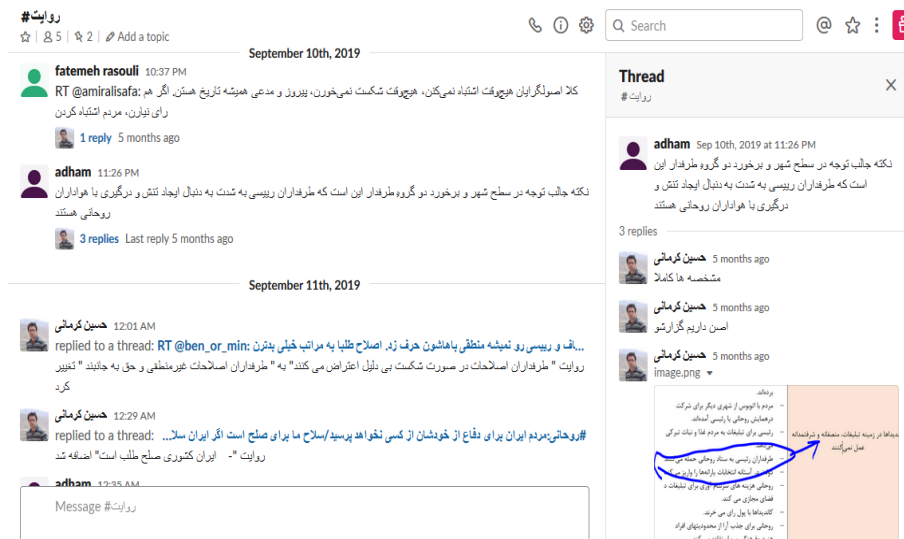


تصویر ۱: شمای کلی کدگذاری توئیت‌ها

هر یک از توئیت‌ها بر اساس شیوه کدگذاری در حداقل دو مرحله منعطف و دایره‌ای در تمام متغیرهای فوق کدگذاری شدند. رهیافت ما در این تحقیق در اینجا نیز تفاوتی با رهیافت‌های موجود در کدگذاری کیفی دارد. در کدگذاری کیفی معمولاً کدگذاران به‌طور جداگانه نمونه‌های متعلق به خود را کدگذاری کرده و سپس با انتخاب درصدی از موارد مشابه، میزان توافق محاسبه می‌گردد (هوپکه، ۲۰۱۵). در این رهیافت بر اساس روش‌های تحلیل ا.ر.ج و مردم‌نگارانه، کار کدگذاری کیفی نیز یک کار جمعی و مشارکتی بود. ما برای این کار از سایت Slack استفاده کردیم. شیوه کار نیز بدین صورت است که هر یک از کدگذاران خودکار کدگذاری را به پیش می‌برد اما در مواردی که شک دارد، موردنظر را ارسال کرده و دیگر کدگذاران نیز نظر خود را درباره آن می‌دهند. در نهایت بر اساس مشارکت همه کدگذاران، کدی که بهترین تناسب را داشته باشد انتخاب می‌شود. این شیوه باعث می‌شود هماهنگی بین کدگذاران بالاتر رفته و همچنین آشنایی آن‌ها با اصول کدگذاری در تحقیق افزایش یابد. در نهایت درصد توافق با توجه به موارد مشابه هر یک از کدگذاران محاسبه می‌گردد.



تصویر ۲، نمونه‌ای از بحث کدگذاران در Slack و همچنین اصلاح و توسعه مقولات در حین کدگذاری نهایی را نشان می‌دهد.



تصویر ۲: کدگذاری مشارکتی در Slack

کدگذاری توئیت‌ها در توئیتر فارسی پیچیدگی زیادی دارد. ما عنوان «پرسش از توئیت» را برای این کار وضع می‌کنیم. در واقع کدگذار در اینجا از توئیت سؤال می‌پرسد و متن (توئیت) پاسخ می‌دهد. در اینجا می‌توان از تمثیل پر کردن پرسشنامه در روش پیمایش استفاده کرد. در روش پیمایش محقق از پاسخگو سؤال می‌پرسد و پاسخگو از بین گزینه‌های موجود یک یا چند گزینه را انتخاب می‌کند. تفاوت عمده در اینجا مشخص نبودن گزینه‌هاست بنابراین گزینه‌های پاسخ هم‌زمان با پاسخگویی شکل می‌گیرند و حتی زمانی که تعدادی گزینه شکل گرفته است (مثلاً در کدگذاری روایت‌ها در این تحقیق، ۹۶ روایت شکل گرفت که هر یک به‌عنوان یک گزینه در نظر گرفته می‌شود)، باز هم پاسخگویی محدود به آن‌ها نیست و گزینه‌ها می‌توانند حذف شده یا گزینه‌های جدید اضافه شوند. همین مسئله نیز به پیچیدگی کدگذاری می‌افزاید. این در حالی است که در

روش‌های کیفی دیگر نیز پس از تعیین برگه کد نهایی، کدگذاری بر اساس آن برگه انجام می‌شود و بنابراین گزینه‌های بالقوه مشخص هستند.

در این تحقیق برای متغیرهایی که در تصویر ۲ معرفی شدند مقولات مشخصی تعریف شد که در تحقیقات دیگر نیز می‌تواند مبنای کار قرار گیرد. غیر از جنس توثیت که توسط نرم‌افزار هم‌زمان با گردآوری داده‌ها مشخص می‌شود و شامل توثیت اصل، منش و ریتوثیت است، مقولات متغیرهای دیگر به‌جز کردارهای شبکه‌ای (قالب‌بندی و روایت‌ها) در جدول ۵ ارائه شده‌اند<sup>۱</sup>. همچنین شخصیت‌ها، با تشخیص نام افراد موجود در توثیت مشخص می‌شود و مقوله‌بندی برای این متغیر لازم نبود.

جدول ۴: مقوله‌بندی سه متغیر تحقیق

جهت‌گیری	لحن توثیت	نوع توثیت
حمایتی	برانگیزاننده	ابراز شخصی
انتقادی	تهییج‌کننده	استفهام انکاری
آگاهی‌بخش	جدی	تفسیر
افشاگرانه	خصمانه	طنز
خنثی	خنثی	خبر
	شاد	سؤال
	شوخ	شعار
	کنایی / طعنه‌آمیز	عبارات غلوآمیز
		گزارش
		مستند
		درخواست سند
		نظرسنجی
		نقل قول مستقیم
		نکته

۱. مقولات این متغیرها در مقالات جداگانه معرفی و تشریح شده‌اند.

جدول ۴ نشان می‌دهد که به‌عنوان مثال، برای تعیین کد مربوط به نوع توثیق در نهایت کد گذار ۱۴ انتخاب داشته است به‌اضافه امکان حذف/تغییر یک مقوله و یا ایجاد مقوله جدید؛ اما عامل دیگری نیز به پیچیدگی کد گذاری اضافه می‌کند. مثلاً تعریف عملیاتی نقل قول در نوع توثیق عبارت است از: نقل قولی که مستقیماً بدون هیچ حذف و اضافه‌ای نقل شده است. همچنین تعریف عملیاتی تفسیر این است: اظهار نظری که به بساخت معنا در مورد یک متن/رویداد می‌پردازد. با توجه به این تعریف‌ها، انتخاب نوع این توثیق دشوار می‌شود: #میرسلیم: سبک زندگی اسلامی خجلی مسئله مهمی است. در دوره دولت شما ما شاهد حقوق‌های نجومی هستیم. این است؟ #مناظره\_دوم. این توثیق یک نقل قول مستقیم از میرسلیم است اما تنها تبدیل واژه «خیلی» به «خجلی» آن را به تفسیر تبدیل می‌کند. همچنین همین تغییر جزئی، لحن توثیق را از جدی به کنایی و جهت‌گیری آن را از خنثی به انتقادی تبدیل می‌کند. باید توجه کرد که اساس نظری این رهیافت که تلفیقی از تحلیل محتوای مردم‌نگارانه و تحلیل گفتمان ا.ر.ج است انعطاف لازم برای در نظر گرفتن این موارد را در اختیار ما قرار می‌دهد.

مورد دیگر که به پیچیدگی کد گذاری توثیق‌ها اضافه می‌کند به قابلیت‌های فنی توثیق برمی‌گردد. در اینجا کد گذاری شخصیت‌ها را مثال می‌زنیم. محقق پیش از انجام تحقیق باید به سؤالات زیر در مورد کد گذاری شخصیت‌ها پاسخ دهد:

- آیا تفاوتی بین نام شخصیت وقتی بدون هشتگ ظاهر می‌شود با وقتی که با هشتگ ظاهر می‌شود وجود دارد؟
- آیا تفاوتی بین ذکر نام و نام خانوادگی شخصیت با ذکر یکی از این دو وجود دارد؟
- آیا هنگامی که عنوان و پست یک شخصیت ذکر می‌شود نیز شخصیت کد گذاری می‌شود یا خیر؟
- آیا شخصیت حقوقی نیز مانند شخصیت حقیقی کد گذاری می‌شود یا خیر؟

- آیا هنگامی که با استعاره/کنایه و یا اشاره غیرمستقیم از یک شخصیت یاد می‌شود هم نام شخصیت کدگذاری می‌شود؟

محقق پیش از کدگذاری شخصیت‌ها باید با توجه به اهداف تحقیق به این پرسش‌ها پاسخ داده تا کدگذاری بر اساس آن‌ها انجام شود. اشتباه یا ابهام در هر یک از مواردی که توضیح دادیم می‌تواند به سوگیری در نتایج منجر شود. مثلاً در توئیت فارسی، از میرحسین موسوی با استعاره «میر دل‌ها» یاد می‌شود و تشخیص اینکه در اینجا نام میرحسین کدگذاری شود یا خیر بر نتایج تحقیق اثر می‌گذارد.

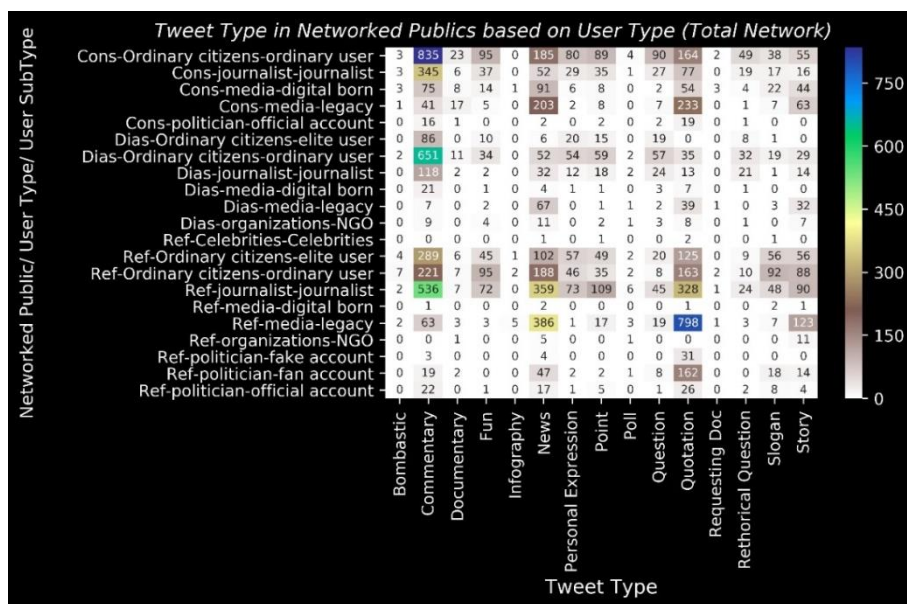
نکته نهایی در این مرحله محاسبه ضریب توافق بین کدگذاران است. ضریب توافق در این رهیافت در دو مرحله، پس از کدگذاری دور اول و کدگذاری نهایی انجام می‌شود. پیشنهاد ما استفاده از ضریب آلفای کریندورف (آلن، ۲۰۱۷) است. لمبارد و دیگران (۲۰۰۲) بیان می‌کنند که ضریب بالای ۸۰ درصد نشان‌دهنده پایا بودن کدگذاری است. تمام ضرایب توافق در این تحقیق بالاتر از این میزان بودند و بنابراین فرایند کدگذاری از قابلیت اعتماد کافی برخوردار بود. در واقع، شیوه کدگذاری مشارکتی در این رهیافت موجب بالارفتن ضریب توافق نیز می‌شود و بدین ترتیب نتایج تحقیق از اعتبار بیشتری برخوردار هستند.

تحلیل‌های آماری و نهایی: پس از کدگذاری کاربران و توئیت‌ها، تحلیل‌های آماری انجام می‌شود. در این مرحله می‌توان از زبان‌های آر و پایتون استفاده کرد. ما در تحقیق خود از نرم‌افزار پایتون استفاده کردیم. در این مرحله محقق می‌تواند بر اساس اهداف و پرسش‌های تحقیق، تقاطع‌های لازم بین متغیرهای مربوط به کاربران و توئیت‌ها را ایجاد کرده و خروجی‌های مختلفی را بگیرد. زبان پایتون ابزار قدرتمندی برای بصری‌سازی خروجی‌ها در اختیار قرار می‌دهد. ما ترجیح دادیم خروجی‌ها را به صورت نقشه حرارتی<sup>۱</sup>

---

1. Heat map

دریافت کنیم. نمونه‌ای از خروجی زبان پایتون که تقاطع بین نوع توئیت و نوع کاربر را بر اساس نوع همگان نشان می‌دهد در تصویر ۳ می‌بینید.



تصویر ۳: نمونه خروجی پایتون (نقشه حرارتی)

همان‌طور که در تصویر ۳ می‌بینیم، زبان پایتون امکان تقاطع هم‌زمان هر تعداد مقوله و زیرمقوله را به محقق می‌دهد و بنابراین محقق می‌تواند با انعطاف بیشتری خروجی‌های موردنظر خود را دریافت کند. پس از گرفتن خروجی‌های موردنیاز کار روش‌شناسی تحقیق به پایان می‌رسد و محقق می‌تواند تحلیل‌های نهایی را به‌منظور پاسخ دادن به پرسش‌های تحقیق آغاز کند. مثلاً تصویر ۳ نشان می‌دهد که بیشترین بسامد مربوط به تفسیرهای کاربران عادی در همگان اصولگرا است. این تقاطع‌ها بر اساس بستر نظری تحقیق می‌تواند تحلیل‌های عمیقی از کنشگری کاربران به دست دهد.

## بحث و جمع‌بندی

هدف این مقاله، ارائه یک رهیافت جامع و دقیق برای انجام تحقیقات اجتماعی در توئیتر بود. اهمیت این کار با توجه به خلأ موجود در این زمینه و اهمیت توئیتر در رویدادهای سیاسی-اجتماعی در جوامع مختلف مشخص می‌شود. تاکنون رهیافتی مشخص و کارآمد که محققان ایرانی بتوانند بر مبنای آن تحقیقات خود در رسانه‌های اجتماعی را هدایت کنند، ارائه نشده است؛ بنابراین ما تلاش کردیم در این تحقیق تا حد امکان این خلأ را رفع کنیم.

رهیافت ارائه‌شده بر مبنای سه روش تحلیل شبکه اجتماعی، تحلیل محتوای مردم نگارانه و تحلیل گفتمان ا.رج شکل گرفته است. تلفیق این سه روش به محقق امکان می‌دهد هم‌زمان کاربر و محتوا را در توئیتر مطالعه کند و بدین ترتیب از محدود شدن تحقیق به یکی از این دو شق اجتناب کند. علاوه بر این، در این رهیافت سه روش به‌طور هم‌زمان بکار گرفته می‌شوند و به همین دلیل دیگر تحقیق صرفاً سویه محاسباتی و یا کیفی نخواهد داشت. همچنین این رهیافت از محدودیت‌های روش‌های مرسوم در تحلیل محتوا و تحلیل گفتمان اجتناب می‌کند. روش‌هایی چون تحلیل محتوای کمی و کیفی و همچنین انواع روش‌های تحلیل گفتمان انتقادی مانند رویکرد فرکلاف، لاکلا-موفه و فوکو به دلیل اینکه مبنای خود را مطالعه و تحلیل متون رسمی قرار داده‌اند، از کارایی کافی برای تحلیل گفتمان رسانه‌های اجتماعی برخوردار نیستند؛ بنابراین ما با مفصل‌بندی تحلیل گفتمان ا.رج با دو روش دیگر تلاش کردیم رهیافتی متناسب با جنس و چستی رسانه‌های اجتماعی ارائه کنیم که بتواند ظرایف و پیچیدگی‌های کنشگری کاربران در این رسانه‌ها را درک و تحلیل کند.

در این مقاله، مراحل مختلف اجرای این رهیافت- از انتخاب بازه زمانی گردآوری داده‌ها تا تحلیل‌های آماری و نهایی- را تشریح کردیم. همچنین با اجرای این رهیافت در

انتخابات ریاست جمهوری سال ۱۳۹۶ به منظور تحلیل کنشگری سیاسی در توئیتر فارسی، کوشیدیم نمونه‌ای عینی از چگونگی اجرای آن ارائه کنیم. در هر مرحله، مفاهیم و ابزارهای مرتبط را معرفی کرده و ارزیابی کردیم تا در نهایت مفاهیم و ابزارهای مورد استفاده در این رهیافت را مشخص سازیم. علاوه بر این، چارچوبی برای شناخت مشخصات جمعیت شناختی کاربران توئیتر (شامل نوع کاربر، هویت، جنسیت و ...) و همچنین چارچوبی کلی برای تحلیل توئیتهای (شامل نوع توئیتهای، لحن توئیتهای، ...) ارائه کردیم. موارد ذکر شده در هر نوع تحقیق روی توئیتر می‌توانند استفاده شوند منتها ما چارچوبی نیز برای تحقیق درباره کنشگری سیاسی در توئیتر فارسی ارائه کردیم که می‌توانند در تحقیقات مشابه مبنایی برای کار محققان فراهم کند.

## منابع

کرمانی، حسین (۱۳۹۸) روایت و شکل‌گیری آن در توئیتر فارسی؛ تحلیل شیوه‌های داستان‌گویی کاربران ایرانی در توئیتر در جریان زلزله کرمانشاه، فصلنامه زبان‌شناسی اجتماعی. دوره ۳، شماره ۱. صص ۵۷-۷۴.

کرمانی، حسین (۱۳۹۵) مشاهده مشارکتی در گروه‌های تلگرامی؛ راهنمای عمل و نمونه موردی، فصلنامه مطالعات فرهنگ-ارتباطات، دوره ۱۷، شماره ۳۵.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). *Fast unfolding of communities in large networks*. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10) <https://doi.org/10.1088/1742-5468/2008/10/P10008>

Borra, E., & Rieder, B. (2014). *Programmed method: developing a toolset for capturing and analyzing tweets*. *Aslib Journal of Information Management*, 66(3), 262–278. <https://doi.org/10.1108/AJIM-09-2013-0094>

Bruns, A., & Stieglitz, S. (2013). *Towards more systematic Twitter analysis: metrics for tweeting activities*. *International Journal of Social Research Methodology*, 16(2), 91–108. <https://doi.org/10.1080/13645579.2012.756095>

Dijck, J. Van, & Poell, T. (2013). *Understanding Social Media Logic*. <https://doi.org/10.12924/mac2013.01010002>

Easley, D., & Kleinberg, J. (2010). *Networks, Crowds, and Markets: A Book by David Easley and Jon Kleinberg*. In *Book*. <https://doi.org/10.1017/CBO9780511761942>

Foucault, Michel (1982) 'The subject and power: An afterword', in Hubert Dreyfus and Paul Rabinow (eds), *Michel Foucault: Beyond Structuralism and Hermeneutics*. Chicago: University of Chicago Press. pp. 208–26.

Golder, Scott A. and Michael Macy. (2012). "Social Science with Social Media." *ASA Footnotes* 40:7.

Himmelboim, I., Sweetser, K. D., Tinkham, S. F., Cameron, K., Danelo, M., & West, K. (2014). *Valence-based homophily on Twitter: Network Analysis of Emotions and Political Talk in the 2012 Presidential Election*. *New Media & Society*, 1461444814555096-. <https://doi.org/10.1177/1461444814555096>

Hooshmand, K. (2015, July). "Soft Power" and its manifestations in international diplomacy. *CGCS Media Wire*. Retrieved from: <http://www.global.asc.upenn.edu/soft-power-and-its-manifestations->



- in-international-diplomacy/  
Jørgensen, M., & Phillips, L. (2002). *Discourse Analysis as Theory and Method*. Thousand Oaks, CA: SAGE Publications.
- Kadushin, C. (2002). *Introduction to Social Network Theory*. Networks. <https://doi.org/10.1007/978-1-4614-2254-9>
- Krippendorff, K. H. (2004). *Content Analysis: An Introduction to Its Methodology (2nd ed.)*. London: Sage.
- Lingel, J. (2017). *Networked Field Studies: Comparative Inquiry and Online Communities*. *Social Media + Society*, 3(4), 205630511774313. <https://doi.org/10.1177/2056305117743139>
- Mahrt, M., & Scharrow, M. (2013). *The Value of Big Data in Digital Media Research*. *Journal of Broadcasting & Electronic Media*, 57(1), 20–33. <https://doi.org/10.1080/08838151.2012.761700>
- Marchant, J., Ormson, T., Honari, A., & Sabeti, A. (2018). *#iranvotes2017: Analysing the 2017 iranian presidential elections through Telegram, Twitter and Instagram*. London.
- Marchant, J., Sabeti, A., Bowen, K., & Kelly, J. (2016). *# IranVotes: Political Discourse on Iranian Twitter during the 2016 Parliamentary Elections*. Retrieved from <https://smallmedia.org.uk/work/iranvotes>
- Marshall, C. (2012). *Big Data, the crowd and me*. *Information Services & Use*, 32(3–4), 215–226. <https://doi.org/10.3233/ISU-2012-0677>
- Nunn, H., & Biressi, A. (2010). 'A trust betrayed': celebrity and the work of emotion. *Celebrity Studies*, 1(1), 49–64. <https://doi.org/10.1080/19392390903519065>
- Parsi, T., Moaveni, A., Jones, M. O., Heyrani, H., & Shahbaz, H. (2018). *Faking the online debate on Iran*. Retrieved September 17, 2018, from Al Jazeera website: <https://www.aljazeera.com/programmes/listeningpost/2018/09/faking-online-debate-iran-180915121527882.html>
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. London: SAGE.
- Schreier, M. (2014). *Qualitative Content Analysis*. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Analysis* (pp. 170–184). London: SAGE Publications.
- Scott, J. (2000). *Social network analysis: A handbook*. In Sage Publications Ltd. <https://doi.org/10.1370/afm.344>
- Snelson, C. L. (2016). *Qualitative and Mixed Methods Social Media Research*. *International Journal of Qualitative Methods*, 15(1), 160940691562457. <https://doi.org/10.1177/1609406915624574>
- Tufekci Z (2014) *Big questions for social media big data: Representativeness, validity and other methodological pitfalls*. In: *Proceedings of the Eighth International AAAI Conference on*

Weblogs and Social Media, Ann Arbor, Michigan, USA, 1–4 June 2014, pp.505–514. Palo Alto, CA: The AAAI Press.