

بررسی کیفیت سوپرگروه‌های تلگرام مبتنی بر آنالیز رفتار کاربران

سید علی هاشمی^۱، دانشجوی دکتری؛ محمدعلی زارع چاهوکی^۲، استادیار

۱- دانشکده مهندسی برق و کامپیوتر - دانشگاه یزد - یزد - ایران

۲- دانشکده مهندسی برق و کامپیوتر - دانشگاه یزد - یزد - ایران

چکیده: امکانات متعددی نظیر کانال، ربات، سوپرگروه، رمزنگاری بدون واسطه و شخصی‌سازی که در پیام‌رسان تلگرام تعبیه شده‌است، محبوبیت این پیام‌رسان را نسبت به بقیه پیام‌رسان‌ها در بین کاربران بیشتر کرده است. امکان ایجاد کانال و سوپرگروه باعث شده‌است که این پیام‌رسان نقش شبکه اجتماعی نیز داشته باشد. با وجود این‌که این پیام‌رسان در حال حاضر بیش از ۲۰۰ میلیون کاربر فعال جهانی و ۴۰ میلیون کاربر فعال ایرانی دارد، پژوهش‌های انگشت شماری بر روی آن انجام شده‌است. در این پژوهش ما رفتار کاربران ایرانی را در سوپرگروه‌های تلگرامی بررسی کرده‌ایم. این پژوهش بر روی بیش از ۳۰۰ هزار سوپرگروه فارسی انجام شده است. نتایج این بررسی درک عمیقی نسبت به محیط تلگرام و رفتار کاربران به خواننده خواهد داد. همچنین، با استفاده از این نتایج معیارهایی برای اندازه‌گیری کیفیت سوپرگروه‌ها در تلگرام تعریف کرده‌ایم. از دیدگاه این مقاله، هر چه اعضای سوپرگروهی به مطالب منتشرشده توجه و علاقه بیشتری داشته باشند، آن سوپرگروه کیفیت بیشتری دارد.

واژه‌های کلیدی: شبکه‌های اجتماعی، پیام‌رسان‌ها، بازاریابی مبتنی بر تلگرام.

An Analytical Study of Users of Telegram Supergroups

S.A. Hashemi¹, PhD Student; M.A. Zare Chahooki², Assistant Professor

1- Faculty of Electrical and Computer Engineering, Yazd University, Yazd, Iran, Email: alihashemi@stu.yazd.ac.ir

2- Faculty of Electrical and Computer Engineering, Yazd University, Yazd, Iran, Email: chahooki@yazd.ac.ir

Abstract: Several features like channels, bots, supergroups, end-to-end encryption, and personalization available in Telegram, has made the instant messenger popular among rivals. The possibility of creating public channels and supergroups has changed Telegram application to a social network rather than an instant messenger. Although Telegram has more than 200 million active worldwide users and 40 million active Iranian users, there is little research based on Telegram. In this paper we have studied Iranian supergroups. More than 300,000 Persian supergroups have been analyzed in our study. The results will give a deep insight of Telegram environment to the reader. We have defined some measures to detect high-quality supergroups, too. In our view, a high-quality supergroup is a supergroup whose members are more engaged with its content.

Keywords: Social networks, Instant messengers, Telegram-based marketing.

تاریخ ارسال مقاله: ۱۳۹۷/۰۷/۱۶

تاریخ اصلاح مقاله: ۱۳۹۷/۱۰/۱۸

تاریخ پذیرش مقاله: ۱۳۹۸/۰۲/۲۰

نام نویسنده مسئول: محمدعلی زارع چاهوکی

نشانی نویسنده مسئول: ایران - یزد - بلوار دانشگاه - دانشگاه یزد - دانشکده مهندسی برق و کامپیوتر.

۱- مقدمه

امروزه برنامه‌های پیام‌رسان تلفن همراه بسیار پرطرفدار هستند. صاحبان این پیام‌رسان‌ها با ارتقای کیفیت و افزودن ویژگی‌های جدید تلاش می‌کنند تا کاربران جدید پیدا کنند و کاربران فعلی خود را از دست ندهند. در حال حاضر تلگرام با حدود ۴۰ میلیون کاربر فعال ایرانی، پرکاربردترین برنامه پیام‌رسان در ایران است. استفاده گسترده از تلگرام در ایران باعث شده‌است که تحقیقاتی در زمینه‌های سیاسی [1]، اجتماعی [2، 3، 4]، گردشگری [5]، رسانه [6، 7، 8]، محیط زیست [9] و زبان‌شناسی [10، 11] مبتنی بر این پیام‌رسان انجام شود. ویژگی‌های خاص تلگرام نظیر کانال، ربات، استیکر آزاد و سوپرگروه باعث شده‌است تا این پیام‌رسان در بین کاربران محبوب شود. سوپرگروه‌های تلگرام می‌توانند تا ۱۰۰۰۰ نفر باشند در حالی که سقف تعداد اعضای گروه‌ها در پیام‌رسان‌های دیگر معمولاً کمتر از ۱۰۰۰ نفر است. هر کاربر تلگرام می‌تواند حداکثر در ۵۰۰ کانال یا سوپرگروه عضو شود. گروه‌ها و کانال‌های عمومی شناسه منحصر به فردی به فرمت *id* دارند. امکان جستجوی گروه‌ها و کانال‌های عمومی با دانستن نام یا شناسه آنها در تلگرام وجود دارد. البته امکان ایجاد و عضویت در گروه‌های عمومی در حال حاضر برای کاربران ایرانی غیرفعال است.

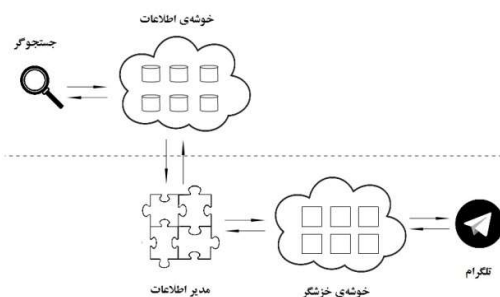
محبوبیت تلگرام در ایران باعث افزایش توجه بازار یاب‌ها به این پیام‌رسان شده است. آقای دانایی و خانم مؤمن [12] تأثیر بازاریابی و بررسی بر قصد خرید مصرف‌کنندگان در تلگرام را بررسی کرده‌اند. آقایان سعیدنیا و قربان‌زاده [13] نیز عوامل مؤثر بر نگرش و تبلیغات شفاهی مثبت در بین کاربران تلگرام را بررسی کرده‌اند. بر خلاف شبکه‌های اجتماعی و پیام‌رسان‌های دیگر امکان انتشار تبلیغات به صورت مکانیزه در تلگرام وجود ندارد. به همین دلیل بازاریاب‌ها برای یافتن کانال یا سوپرگروه مناسب تبلیغاتشان مجبور به جستجو و بررسی چشمی هستند. این موضوع باعث می‌شود که انتشار تبلیغات هدفمند در تلگرام برای بازاریاب‌ها کاری دشوار و زمان‌بر باشد. به همین دلیل برخی از بازاریاب‌ها ترجیح می‌دهند با کاهش هدفمندی تبلیغات، یعنی انتشار آن در کانال‌ها و سوپرگروه‌های پرمخاطب به جای کانال‌ها و سوپرگروه‌های مرتبط، کار خود را ساده‌تر کنند. این کار به نفع تعداد اندکی کانال و سوپرگروه پرمخاطب است ولی به ضرر بازاریاب‌ها، کانال‌ها و سوپرگروه‌های دیگر و همچنین کاربران تمام خواهد شد. کاربران، تبلیغات متعدد و نامرتب با علائق‌شان مشاهده خواهند کرد که باعث نارضایتی آن‌ها خواهد شد. کانال‌ها و سوپرگروه‌هایی که در بین پرمخاطب‌ترین‌ها نیستند، درآمد تبلیغاتی نخواهند داشت که باعث کاهش انگیزه مدیران آن‌ها برای ارائه محتوای باکیفیت خواهد شد. بازاریاب‌ها نیز به دلیل هدفمند نبودن تبلیغاتشان، بازدهی و نرخ تبدیل کمتری تجربه خواهند کرد.

مشکل دیگری که بر سر راه بازاریاب‌ها قرار دارد عدم وجود معیار مناسب و قاطعیت کافی در تعیین هزینه تبلیغات و بررسی کیفیت کانال‌ها و سوپرگروه‌ها است. اعضای کانال در تلگرام فقط برای مدیران

کانال قابل مشاهده است. برخی از کانال‌ها با افزودن کاربران غیرواقعی، سعی در تقلب در آمار اعضا و بازدید دارند تا بتوانند هزینه بیشتر از تبلیغ‌دهندگان دریافت کنند. هزینه تبلیغات معمولاً از طرف صاحب کانال یا سوپرگروه به صورت سلیقه‌ای تعیین می‌شود. عدم وجود ابزارهای لازم باعث شده‌است که تخمین اثربخشی و بازدهی تبلیغات بسیار دشوار باشد. برخی از بازاریاب‌ها بازدهی تبلیغات را با سعی و خطا به دست می‌آورند. آن‌ها نرخ تبدیل و بازدهی را پس از یک بار انتشار تبلیغات محاسبه می‌کنند و در صورت رضایت تبلیغات خود را مجدداً در همان کانال یا سوپرگروه منتشر می‌کنند. انواع تبلیغات تلگرامی توسط آقای نوبری و همکاران [14] با آنالیز گراف بررسی شده است.

مشکلات مذکور باعث شد تا سامانه ایده‌کاو را ایجاد کنیم. هدف از این سامانه بهبود کیفیت و تسهیل بازاریابی در محیط تلگرام است. امکان جستجوی محتوای سوپرگروه‌ها و کانال‌ها در ایده‌کاو فراهم شده است. الگوریتم رتبه‌بندی ایده‌کاو، نتایج مرتبط و باکیفیت را در بالای لیست نمایش می‌دهد. همچنین سوپرگروه‌های مشابه با نتایج جستجو، با توصیه‌گر ایده‌کاو به بازاریاب‌ها پیشنهاد می‌شوند. جزئیات طراحی و معماری سامانه ایده‌کاو را در بخش ۲ بررسی خواهیم کرد. ایده‌کاو انبوهی از سوپرگروه‌ها و کانال‌های تلگرامی را به صورت لحظه‌ای خزش و رصد می‌کند. در این مقاله اطلاعات به دست آمده از بیش از ۳۰۰ هزار سوپرگروه تلگرام را برای اولین بار بررسی می‌کنیم. تا آنجا که ما مطلع هستیم، تا به حال هیچ پژوهشی بر روی تلگرام با این گستردگی جامعه آماری انجام نشده است. هدف نهایی بررسی‌های ما، معرفی معیارهایی برای تعیین کیفیت سوپرگروه‌های تلگرام است. از این معیارها در ایده‌کاو برای رتبه‌بندی سوپرگروه‌ها استفاده می‌شود.

پس از این مقدمه، در بخش ۲ سامانه ایده‌کاو را معرفی کرده و معماری آن را بررسی خواهیم کرد. در بخش ۳ نحوه جمع‌آوری داده و ایجاد دادگان را توضیح خواهیم داد. در بخش ۴ نتایج اولیه را بررسی می‌کنیم. در بخش ۵ وارد جزئیات بیشتری خواهیم شد و از زوایای دیگری به نتایج نگاه می‌کنیم. بخش ۶ نیز به نتیجه‌گیری و ایده‌هایی برای پژوهش‌های آتی اختصاص یافته است.



شکل ۱ شمای کلی معماری ایده‌کاو

۲- سامانه ایده‌کاو

در نگاه کلی، ایده‌کاو از سه بخش خزش‌گر^۱، مدیر اطلاعات و جستجوگر تشکیل شده‌است. خزش‌گر با شنود گروه‌های تلگرامی، داده‌های آن‌ها را به صورت لحظه‌ای ذخیره می‌کند. داده‌های خزش‌گر برای پردازش به بخش مدیر اطلاعات ارسال می‌شود. این داده‌ها به صورت آنلاین پردازش شده و اطلاعات مستخرج از آن به شکل مناسبی ذخیره‌سازی می‌شود. جستجوگر ایده‌کاو این اطلاعات را جستجو و رتبه‌بندی می‌کند. شکل ۱ شمای کلی معماری ایده‌کاو را نمایش می‌دهد.

خزش‌گر ایده‌کاو بر پایه پروتکل MTPROTO به زبان پایتون پیاده‌سازی شده‌است. هر شماره تلفن یک نمونه^۲ از خزش‌گر را اجرا می‌کند. نمونه‌های خزش‌گر در کنار هم استخراج خزش‌گر را ایجاد می‌کنند. استخراج خزش‌گر با برنامه‌ریز^۳ هوشمند، اجرای نمونه‌های خزش‌گر را مدیریت می‌کند. هر نمونه خزش‌گر با تخصیص حافظه اندکی قابل اجراست. استخراج خزش‌گر مقاوم در برابر خطا^۵ است. خطای نمونه‌ها باعث از کار افتادن استخراج نمی‌شود.

خزش‌گر ایده‌کاو همانند خزش‌گرهای تحت وب، از تعدادی گروه هسته^۶ شروع به کار کرده و با دنبال کردن لینک‌های جدید عضو گروه‌های تازه می‌شود. پس از عضویت در هر گروه، امکان خواندن پیام‌ها، اعضا و اطلاعات تکمیلی برای خزش‌گر فراهم می‌شود. هر نمونه از خزش‌گر دائماً در حال رصد کردن تمامی گروه‌هایی است که در آن‌ها عضو است. رویدادهایی^۷ مانند پیام جدید و عضو جدید به صورت لحظه‌ای از تلگرام دریافت می‌شود. همه رویدادها توسط خزش‌گر ایده‌کاو ثبت می‌شوند.

داده‌های به‌دست‌آمده از خزش‌گر نظیر پیام‌ها و اطلاعات عضویت افراد در گروه‌ها به بخش مدیر اطلاعات ارسال می‌شود. مدیر اطلاعات از این داده‌ها، اطلاعات مورد نیاز را استخراج کرده و با ساختمان داده‌ای مناسب ذخیره‌سازی می‌کند. لینک‌های جدید در این بخش پیدا شده و برای عضویت به خزش‌گر ارسال می‌شود. محتوای گروه‌ها نیز همین‌جا بررسی می‌شود. گروه‌هایی که محتوای اسپم یا کم‌ارزش دارند شناسایی شده و برای ترک کردن به خزش‌گر فرستاده می‌شود. اطلاعات مورد نیاز دیگر نیز در این بخش نمایه‌سازی می‌شوند. ذخیره‌سازی اطلاعات بلادرنگ بوده و به صورت توزیع‌شده بین سرورهای خوشه^۸ انجام می‌شود. ذخیره‌سازی در تکه‌های^۹ کوچک به تعداد زیاد انجام می‌شود.

جستجوگر ایده‌کاو بر پایه نمایه‌ساز^{۱۰} و جستجوگر لوسین^{۱۱} پیاده‌سازی شده‌است. جستجو بر روی سه نمایه گروه‌ها، کاربران و پیام‌ها قابل انجام است. معماری ذخیره‌سازی اطلاعات، نمایه معکوس^{۱۲} است. نمایه معکوس سرعت بالایی در بازیابی اطلاعات دارد. علاوه بر این، امکان اعمال فیلترهای مختلف بر روی نمایه معکوس وجود دارد. معماری جستجوی اطلاعات، TFIDF است. در این معماری کلمات مهم متن به سرعت پیدا شده و از کلمات عمومی متمایز می‌شوند.

جستجو با ارسال جستار^{۱۳} می‌تواند نتایج مرتبط با موضوع مورد نظر کاربر را بازیابی کند. عموماً در این نوع جستار، نتایجی که در رتبه بالا

قرار دارند بسیار مرتبط هستند ولی هر چه پایین‌تر برویم ارتباط جستار با اسناد بازیابی شده کمتر می‌شود. معماری اصلی مورد استفاده برای جستجوی با جستار، معماری نمایه‌ای است. در این معماری بازیابی نتایج مرتبط سریع است ولی هرچه پایین‌تر برویم بازیابی زمان بیشتری می‌گیرد. این مشکلات باعث می‌شود جستجوی مبتنی بر جستار به تنهایی نتواند جوامع هدف بزرگ را بازیابی کند.

۲-۱- معماری ایده‌کاو

همان‌طور که ذکر شد، جستجوی ایده‌کاو بر روی نمایه معکوس توزیع‌شده انجام می‌شود. اطلاعات متنی گروه‌های تلگرامی شامل عنوان، درباره و ده‌هزار آخرین پیام ارسال‌شده در گروه نمایه‌سازی می‌شوند. وزن امتیاز TFIDF عنوان گروه برابر ۵، متن درباره گروه ۲ و پیام‌های گروه یک است. جستجوی متنی کاربران ایده‌کاو بر روی همین نمایه توزیع‌شده انجام می‌شود.

فرایند ایجاد و به‌روزرسانی نمایه توزیع‌شده توسط مدیر اطلاعات ایده‌کاو انجام می‌شود. مدیر اطلاعات عنوان، درباره و آخرین پیام‌های گروه‌ها را به صورت لحظه‌ای از خزش‌گر دریافت می‌کند. خزش‌گر ایده‌کاو مانند خزش‌گرهای بر بستر وب، وارد هر گروهی که می‌شود لینک‌های موجود در آن گروه توسط خزش‌گر خوانده می‌شود. لینک‌ها به مدیر اطلاعات ارسال می‌شوند. مدیر اطلاعات لینک‌ها را با توجه به کیفیت لینک‌های گروه‌های مبدأ و تعداد دفعاتی که توسط خزش‌گر مشاهده شده‌اند، اولویت‌بندی می‌کند. خزش‌گر به صورت دوره‌ای لینک‌ها را با توجه به اولویت آن‌ها از مدیر اطلاعات دریافت کرده و اقدام به عضویت در گروه‌های جدید می‌کند.

فرایند فوق‌الذکر باعث می‌شود که خزش‌گر دائماً گروه‌های جدید را شناسایی کند و در آن‌ها عضو شود. البته برخلاف وب، در تلگرام محدودیت‌هایی در این زمینه وجود دارد. محدودیت عضویت در حداکثر ۵۰۰ گروه برای هر شماره تلفن توسط تلگرام اعمال شده‌است. به همین دلیل خزش‌گر ایده‌کاو به صورت توزیع‌شده برنامه‌ریزی شده‌است. خوشه خزش‌گر ایده‌کاو شامل صدها نمونه اجرایی است که هر کدام با شماره تلفن مجزا اجرا می‌شوند.

خزش‌گر ایده‌کاو علاوه بر عضویت در گروه‌ها و خواندن پیام‌های آن‌ها، اطلاعات تمام اعضای گروه را نیز استخراج می‌کند. مدیر اطلاعات علاوه بر اطلاعات گروه‌ها، اطلاعات کاربران را نیز ذخیره می‌کند. با پیوند دادن اطلاعات گروه‌ها و کاربران، مدیر اطلاعات می‌داند که هر کاربر در کدام گروه‌ها عضو یا مدیر است.

تمامی پیام‌های هر گروه توسط خزش‌گر خوانده شده و برای مدیر اطلاعات ارسال می‌شود. مدیر اطلاعات هر یک از پیام‌ها را هم از نظر محتوا (مثلاً تعداد لینک‌ها و شماره تلفن‌ها) و هم از نظر ارتباطات (نظیر فروردی‌بودن و پاسخ پیام دیگری بودن) پردازش و بررسی می‌کند. این اطلاعات نیز در نمایه ذخیره‌سازی می‌شود.

۳- جمع‌آوری داده و ایجاد دادگان

داده‌های بررسی شده در این مطالعه، اطلاعات خزش شده از تلگرام در یک بازه دو ماهه است. این اطلاعات در نمایه‌ای مجزا برای آنالیز ذخیره‌سازی شده‌است.

جدول ۱ پارامترهای ذخیره‌شده در این نمایه را نشان می‌دهد. به‌جز پارامترهای ۳، ۴، ۱۷، ۱۸ و ۱۹ در

جدول ۱ مقدار همه پارامترها بین سوپرگروه‌ها قابل‌مقایسه است.

جدول ۱ پارامترهای نمایه

شناسه	پارامتر	توضیحات
۱	AUC	تعداد کاربرانی که حداقل یک پیام در گروه ارسال کرده‌اند
۲	AC	تعداد مدیران گروه
۳	T	عنوان گروه
۴	DT	تاریخ پیداشدن گروه توسط ایده‌کاو
۵	AL	میانگین طول پیام‌های گروه
۶	CA	میانگین تعداد فروردهایی که از هر کانال در گروه منتشر شده‌است
۷	CC	تعداد کانال‌های منحصربه‌فردی که حداقل یک پیام از آن‌ها در گروه فرورده شده‌است
۸	CV	واریانس تعداد پیام‌هایی که از هر کانال در گروه فرورده شده‌است
۹	CH	ساعات تعطیلی گروه (بین ۰ تا ۲۴)، تعداد ساعاتی از شبانه‌روز که در هر تاریخی هیچ پیامی در گروه ارسال نشده‌است
۱۰	FC	تعداد پیام‌هایی که در گروه فرورده شده‌است
۱۱	GNL	تعداد کاراکترهای موجود در نام گروه
۱۲	P	تعداد اعضای گروه
۱۳	DC	تعداد پیام‌های منحصربه‌فردی که بیش از یک بار در گروه ارسال شده‌اند (تکراری هستند)
۱۴	DS	مجموع تعداد پیام‌های تکراری گروه
۱۵	LC	تعداد لینک‌های منتشرشده در گروه
۱۶	MC	تعداد پیام‌های چندرسانه‌ای (شامل عکس، فیلم و...) گروه
۱۷	A	متن درباره گروه
۱۸	MC	متن تمامی پیام‌های گروه
۱۹	PI	شناسه منحصربه‌فرد گروه
۲۰	PC	تعداد شماره‌تلفن‌های منتشرشده در گروه
۲۱	RC	تعداد پیام‌هایی که در پاسخ به پیام دیگری در گروه آمده‌است
۲۲	TC	تعداد کل پیام‌های منتشرشده در گروه
۲۳	WCN	تعداد کاراکترهای نام‌تعارف (غیر از اعداد و حروف) در نام گروه

۴- نتایج بررسی‌های اولیه

در این بخش نتایج بررسی‌های اولیه بر روی دادگان و مقایسه آن‌ها با دیگر داده‌ها را آورده‌ایم. برای هر یک از پارامترهای

جدول ۱ بین سوپرگروه‌ها مقایسه انجام شده‌است. همچنین بین هر دو پارامتر نیز نرخ و نسبت رشد محاسبه و بین سوپرگروه‌ها مقایسه شده‌است. نتایج مهم و قابل‌توجه این بررسی‌ها در ادامه آمده است.

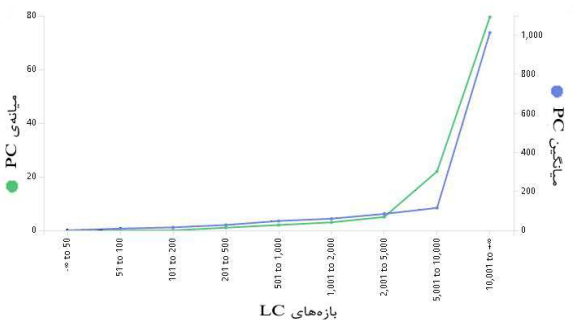
طول پیام‌های منتشرشده در گروه‌های تخصصی به‌شکل قابل‌توجهی بیشتر از بقیه گروه‌ها است. در دادگان ما، میانگین تعداد کاراکترهای هر پیام در گروه‌های تخصصی ۱۳۶ است، در صورتی که این مقدار برای گروه‌های چت‌وگفتگو ۴۸ است. طبیعتاً گروه‌های چت‌وگفتگو با توجه به ماهیت شان پیام‌های کوتاه (نظیر سلام و احوال‌پرسی) زیادی ارسال می‌کنند که در گروه‌های دیگر رایج نیست. گروه‌های تخصصی نیز تمایل انتشار یا بازنشر مطالب علمی و تخصصی دارند که غالباً طولانی هستند. بنابراین این آمار با ماهیت این گروه‌ها مطابقت دارد.

گروه‌های تخصصی تنوع بیشتری در پیام‌های فرورده‌ای نسبت به گروه‌های چت‌وگفتگو دارند. در هر گروه تخصصی مورد بررسی به‌طور میانگین از ۳۶ کانال مختلف پیام فرورده شده‌است (میانگین: ۲۰)، در صورتی که در گروه‌های چت‌وگفتگو به‌طور میانگین از ۲۰ کانال مختلف پیام فرورده شده‌است (میانگین: ۷). میانگین تعداد پیام‌های فرورده‌ای در گروه‌های تخصصی ۵۱ است. در گروه‌های چت‌وگفتگو میانگین تعداد پیام‌های فرورده‌ای ۱۷ است. پس گروه‌های تخصصی نسبت به گروه‌های چت‌وگفتگو تمایل بیشتری به فرورده‌کردن پیام‌های دیگران دارند.

گروه‌های چت‌وگفتگوی مورد بررسی به‌طور میانگین ۴۶۸۹ پیام در پاسخ به پیام دیگری ارسال کرده‌بودند (میانگین: ۹۶۱). این عدد برای گروه‌های تخصصی ۴۲۷ بود (میانگین: ۲۰۱). با توجه به این که طبیعت گروه‌های چت‌وگفتگو پرسش‌وپاسخ زیاد است، این نتیجه کاملاً قابل‌توجه است. البته زمانی که نرخ پاسخ‌ها به تعداد کل پیام‌ها را محاسبه می‌کنیم، این نتیجه کمتر به چشم می‌آید. به‌طور میانگین، در گروه‌های چت‌وگفتگو ۴۸٪ از پیام‌ها، پاسخ پیام دیگری بوده‌است. در گروه‌های تخصصی ۳۵٪ از پیام‌ها پاسخ بوده‌است، درحالی‌که به‌طور کلی در بین همه گروه‌ها نرخ پاسخ ۲۰٪ است. این آمار نشان می‌دهد که گروه‌های تخصصی نسبت به بقیه گروه‌ها بیشتر به پرسش‌وپاسخ می‌پردازند، ولی میزان این پرسش‌وپاسخ به اندازه گروه‌های چت‌وگفتگو نیست. همچنین، گروه‌های چت‌وگفتگو تعداد پیام‌های بیشتری را نسبت به گروه‌های تخصصی ارسال می‌کنند.

۵- بحث و بررسی بیشتر

تعداد لینک‌های منتشرشده در گروه‌های تلگرامی از تعداد شماره تلفن‌ها با اختلاف قابل توجهی بیشتر است، ولی تعداد لینک‌ها با تعداد شماره تلفن‌ها در هر گروه متناسب هستند. به طور میانگین در هر گروه تلگرام تعداد لینک‌های منتشرشده ۱۳۱ برابر تعداد شماره تلفن‌ها است. میان نسبت تعداد لینک‌ها به شماره تلفن‌ها ۱۵ است. در گروه‌های چت‌وگفتگوی دادگان ما این میانگین ۵۰ و میان ۱۵ است. در گروه‌های تخصصی میانگین ۴۰ و میان ۳۶ است. بنابراین در گروه‌های تخصصی شماره تلفن‌های بیشتری نسبت به بقیه گروه‌ها منتشر می‌شود. علت این افزایش این است که ارائه‌دهندگان خدمات در گروه‌های تخصصی شماره تلفن تماس خود را نیز به اشتراک می‌گذارند. در گروه‌های غیرتخصصی کمتر خدمات با شماره تلفن تماس معرفی می‌شوند. شکل ۲ میان و میانگین تعداد شماره تلفن‌های منتشرشده در گروه‌های تلگرامی را در بازه‌های مختلف تعداد لینک‌ها نشان می‌دهد. همان‌طور که مشاهده می‌شود، روند رشد میان و میانگین تعداد شماره تلفن‌ها با روند رشد تعداد لینک‌ها متناسب است. بر همین اساس می‌توان گفت که تعداد لینک‌ها و تعداد شماره تلفن‌های منتشرشده در گروه‌های تلگرامی با هم رابطه مستقیم دارند.



شکل ۲ رشد تعداد شماره تلفن‌ها با افزایش تعداد لینک‌ها

به‌طور کلی گروه‌هایی که کاراکترهای غیرعادی بیشتری در نام خود دارند، اعضای بیشتری نیز دارند. البته اکثر گروه‌هایی که کاراکترهای غیرعادی زیادی در نام خود دارند، دارای محتوای کم‌ارزش برای جستجوی متنی هستند. چند نمونه از این گروه‌ها عبارت‌اند از: "چت‌وگفتگوی دادگان"، "دانشگاه تبریز"، "گروه‌های تخصصی"، "گروه‌های غیرتخصصی"، "گروه‌های چت‌وگفتگوی دادگان"، "گروه‌های غیرتخصصی"، "گروه‌های چت‌وگفتگوی دادگان"، "گروه‌های غیرتخصصی"، "گروه‌های چت‌وگفتگوی دادگان"، "گروه‌های غیرتخصصی"، "گروه‌های چت‌وگفتگوی دادگان"، "گروه‌های غیرتخصصی". کاراکترهای غیرعادی عموماً برای منحصربه‌فرد کردن یا زیباتر شدن نام گروه استفاده می‌شود. آمار شکل ۳ نشان می‌دهد که کاربران تمایل بیشتری به عضویت در این گروه‌ها دارند. البته با بررسی دقیق‌تر پیام‌ها و تعاملات گروه‌هایی که در نام خود کاراکترهای غیرعادی زیادی دارند، به این نتیجه می‌رسیم که اکثریت اعضای این گروه‌ها توسط اعضای دیگر به گروه اضافه شده‌اند. اکثر این گروه‌ها به ازای عضوکردن دیگران به اعضای خود پاداش‌هایی نظیر شارژ رایگان یا امکان ارسال پیام تبلیغاتی می‌دهند. این پاداش‌ها، اعضای فعلی را به عضوکردن دیگران ترغیب می‌کند.

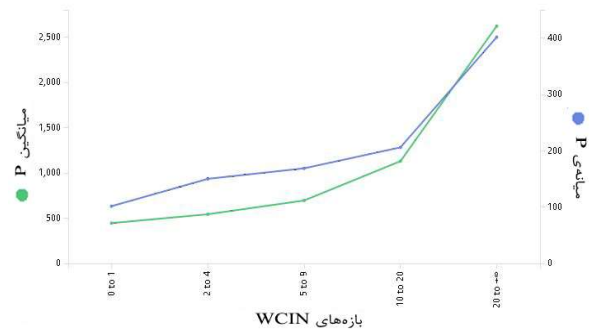
گروه‌های چت‌وگفتگوی مورد بررسی ۸۴۹۶ پیام به‌طور میانگین (در بازه زمانی ۲ ماهه) ارسال کرده‌بودند. میانگین پیام‌های گروه‌های تخصصی ۱۴۴۵ بود که دقیقاً برابر با میانگین پیام‌های همه گروه‌ها به طور کلی است. گروه‌های چت‌وگفتگوی جزو فعال‌ترین گروه‌های تلگرامی از لحاظ تعداد پیام‌های ارسالی هستند. البته زمانی نسبت تعداد کاربران فعال به تعداد پیام‌ها را محاسبه می‌کنیم، متوجه می‌شویم که میزان درگیری کاربران گروه‌های تخصصی بیشتر از گروه‌های چت‌وگفتگوی بوده‌است. ۱۴۴۵ پیام در گروه‌های تخصصی به طور میانگین توسط ۲۰۵ کاربر ارسال شده‌است، درحالی‌که ۸۴۹۶ پیام در گروه‌های چت‌وگفتگوی توسط ۵۷۰ کاربر ارسال شده‌است. بنابراین میزان درگیری کاربران گروه‌های تخصصی ۱۷٪ و گروه‌های چت‌وگفتگوی ۷٪ محاسبه می‌شود. این آمار نشان می‌دهد که در گروه‌های تخصصی افراد بیشتری در بحث‌ها شرکت می‌کنند ولی بحث‌ها کوتاه است. در گروه‌های چت‌وگفتگوی افراد کمتری در بحث‌ها مشارکت می‌کنند ولی بحث‌ها طولانی‌تر است.

گروه‌های تخصصی عموماً از کاراکترهایی غیر از حروف و اعداد در نام گروه استفاده نمی‌کنند. گروه‌های چت‌وگفتگوی برخلاف آن‌ها، از این کاراکترها زیاد استفاده می‌کنند. میان استفاده از کاراکترهای غیرمعمول در نام گروه برای گروه‌های تخصصی ۲ و برای گروه‌های چت‌وگفتگوی ۹ است.

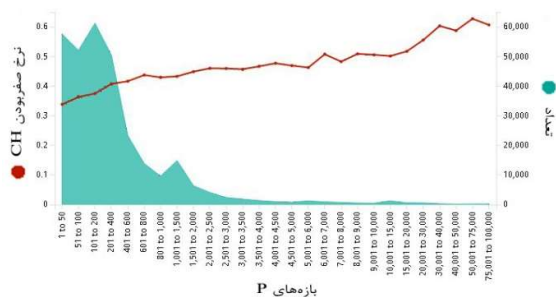
علاوه بر موارد مذکور، به‌طور کلی با بررسی نسبت دوبره‌دوی پارامترها با ضریب اطمینان بالایی به نتایج زیر دست یافتیم:

- ۱- نسبت میانگین طول پیام‌ها به مجموع پیام‌های تکراری هر چه بیشتر باشد بهتر است.
- ۲- نسبت میانگین طول پیام‌ها (شناسه ۵ در جدول ۱) به تعداد لینک‌ها (شناسه ۱۵ در جدول ۱) هر چه بیشتر باشد بهتر است.
- ۳- نسبت تعداد پیام‌های فروردی (شناسه ۱۰ در جدول ۱) به تعداد پیام‌های منحصربه‌فرد تکراری (شناسه ۱۳ در جدول ۱) هر چه بیشتر باشد بهتر است.
- ۴- نسبت تعداد پیام‌های فروردی (شناسه ۱۰ در جدول ۱) به تعداد پیام‌های تکراری (شناسه ۱۳ در جدول ۱) هر چه بیشتر باشد بهتر است.
- ۵- نسبت تعداد پیام‌های فروردی (شناسه ۱۰ در جدول ۱) به مجموع پیام‌های تکراری (شناسه ۱۴ در جدول ۱) هر چه بیشتر باشد بهتر است.
- ۶- نسبت تعداد اعضا (شناسه ۱۲ در جدول ۱) به تعداد کل پیام‌ها (شناسه ۲۲ در جدول ۱) هر چه بیشتر باشد بهتر است.
- ۷- نسبت تعداد مدیرها (شناسه ۲ در جدول ۱) به تعداد کاراکترهای غیرعادی (شناسه ۲۳ در جدول ۱) هر چه بیشتر باشد بهتر است.

می‌دهد که ساعات خاموشی ندارند. همان طور که مشاهده می‌شود از بین حدود ۱۵۰ هزار گروهی که کمتر از ۲۰۰ نفر عضو دارند، تقریباً ۳۵٪ از آن‌ها ساعات خاموشی ندارند. هر چه تعداد اعضای گروه‌ها زیاد می‌شود، درصد گروه‌های بدون ساعات خاموشی نیز تا حدود ۶۰٪ افزایش می‌یابد. میانگین تعداد اعضای گروه‌های تخصصی ۲۹۹۰ نفر و میانگین تعداد اعضای چت‌وگفتگو ۱۲۱۵۰ نفر است. درحالی که میانگین تعداد اعضا به طور کلی ۷۷۸ نفر است. روند کاهش ساعات خاموشی با افزایش اعضای گروه، طبیعی به نظر می‌آید. هر چه اعضای گروهی بیشتر باشند، احتمال ارسال پیام‌های نامنظم در آن گروه بیشتر می‌شود.



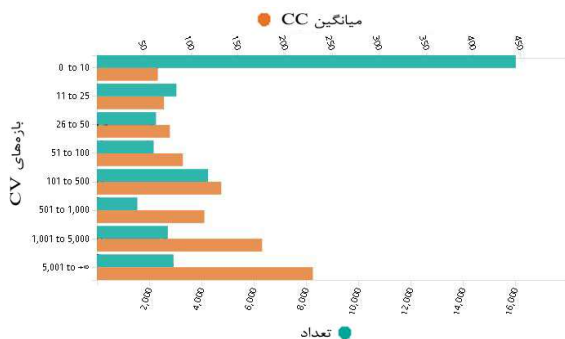
شکل ۳ رشد تعداد اعضا با افزایش کاراکترهای غیرعادی



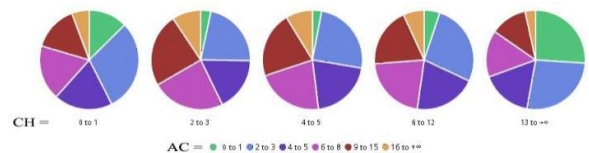
شکل ۵ نرخ گروه‌های بدون ساعات خاموشی (قرمز) به همراه تعداد گروه‌های هر بازه اعضا (سبز)

اکثر گروه‌ها پیام‌ها را با واریانس خیلی کمی فرورد می‌کنند. به عبارت دیگر توزیع پیام‌های کانال‌های مختلف در گروه یکسان است. نمودار شکل ۶ فقط برای گروه‌هایی که حداقل از ۲ کانال مختلف در آن‌ها پیام فرورد شده‌است، رسم شده‌است. بیش از ۵۰٪ از این گروه‌ها در اولین بازه یعنی واریانس کمتر از ۱۰ قرار گرفته‌اند. به‌طور میانگین هر یک از این گروه‌ها از ۱۰۰ کانال مختلف پیام فرورد کرده‌اند. تعداد پیام‌های فروردی از هر کانال به قدری به هم نزدیک بوده‌است که واریانس به‌دست‌آمده برای این گروه‌ها از ۱۰ کمتر شده‌است. در طرف مخالف، درصد اندکی از گروه‌ها به‌طور میانگین از بیش از ۴۵۰ کانال مختلف پیام فرورد کرده‌اند.

ساعات خاموشی گروه‌ها با افزایش تعداد مدیران گروه، کاهش می‌یابد. در شکل ۴، ۵۰٪ از گروه‌هایی که ساعات خاموشی آن‌ها کمتر یک ساعت بوده‌است، بیشتر از ۵ مدیر داشته‌اند. در صورتی که گروه‌های با تعداد مدیر بیش از ۵ فقط ۳۰٪ از گروه‌های دارای بیش از ۱۳ ساعت خاموشی را شامل می‌شوند. گروه‌هایی که بین ۳ تا ۸ مدیر داشته‌اند در همه حالات ساعات خاموشی حدود ۵۰٪ از آمار را به خود اختصاص داده‌اند. میانگین تعداد مدیر گروه در بین همه گروه‌ها، ۵ مدیر و میانگین تعداد مدیر ۳ است. رفتار گروه‌های تخصصی و چت‌وگفتگو در این معیار مشابه بوده و هر دو به‌طور میانگین ۱۰ مدیر دارند که میانگین آن ۱۷ است. میانگین ساعات خاموشی در گروه‌های تخصصی دو ساعت و نیم و در گروه‌های چت‌وگفتگو یک ساعت و نیم است. بنابراین تعداد مدیران و ساعات خاموشی معیار مناسبی برای تمایز قائل شدن بین این دو نوع گروه نیست. البته ساعات خاموشی را می‌توان معیاری برای فعالیت منظم گروه‌ها در نظر گرفت. گروه‌هایی که ساعات خاموشی کمتری دارند، نظم فعالیت کمتری دارند. کاربران این گروه‌ها در هر ساعت از شبانه‌روز پیام ارسال می‌کنند که احتمال خوانده شدن پیام‌های گروه توسط اعضا را کاهش می‌دهد. در نقطه مخالف، گروه‌هایی که ساعات خاموشی زیادی دارند، فعالیتشان منظم است. به طور مثال یک گروه مشاوره تخصصی، فقط در بازه مشخصی از ساعات روز فعالیت می‌کند. با توجه به این که تمامی فعالیت گروه در آن بازه نسبتاً کوتاه انجام می‌شود، احتمال خوانده شدن پیام‌ها و تعامل اعضای گروه با آن‌ها افزایش می‌یابد.



شکل ۶ تعداد گروه‌ها و میانگین تعداد کانال فروردشده در بازه‌های مختلف واریانس کانال‌ها

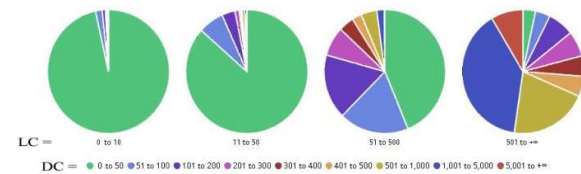


شکل ۴ نمودار تعداد مدیران گروه بر اساس ساعات خاموشی

البته تأثیر تعداد مدیران گروه به اندازه تأثیر تعداد اعضای گروه نیست. در شکل ۵ نموداری بر اساس تعداد اعضای گروه‌ها رسم شده‌است. ناحیه سبز نشان‌دهنده تعداد کل گروه‌های موجود در آن بازه اعضا است. خط قرمز رنگ درصد گروه‌هایی در آن بازه را نشان

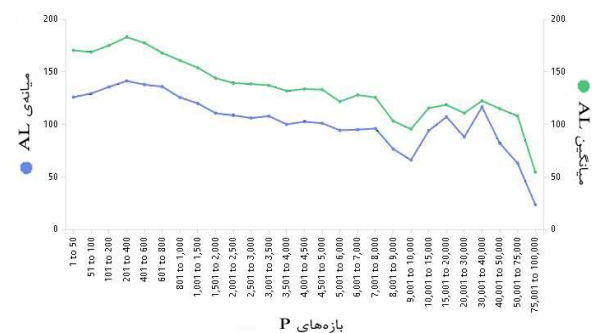
تعداد پیام‌های فروراد شده از برخی از این کانال‌ها خیلی بیشتر از برخی دیگر بوده‌است، که باعث شده تا واریانس بالا برود. بین گروه‌های چت‌وگفتگو و گروه‌های تخصصی در این نمودار تفاوت چشم‌گیری مشاهده نشد. یکی از دلایل عمده واریانس بالای پیام‌های فرورادی، تبلیغات یک یا چند کانال خاص در یک گروه می‌باشد. پیام‌های کانال تبلیغ‌دهنده به تعداد زیاد در گروه فروراد می‌شود که باعث افزایش چشم‌گیر واریانس پیام‌های فرورادی است.

روند رشد تعداد پیام‌های تکراری گروه با تعداد لینک‌های منتشر شده در گروه مشابه است. این روند برای هر دو دسته از گروه‌های چت‌وگفتگو و تخصصی نیز مشابه است. با توجه به آمار نمایش داده شده در شکل ۷ به نظر می‌آید که اکثر پیام‌های تکراری حاوی لینک هستند. به عبارت دیگر هر چه گروهی بیشتر به انتشار لینک تمایل داشته باشد، پیام تکراری بیشتری خواهد داشت. دلیل عمده این روند نیز تبلیغات است. پیام‌های تبلیغاتی شامل لینک هستند. برای بیشتر مشاهده شدن تبلیغات، هر پیام به تعداد زیاد در گروه تکرار می‌شود.



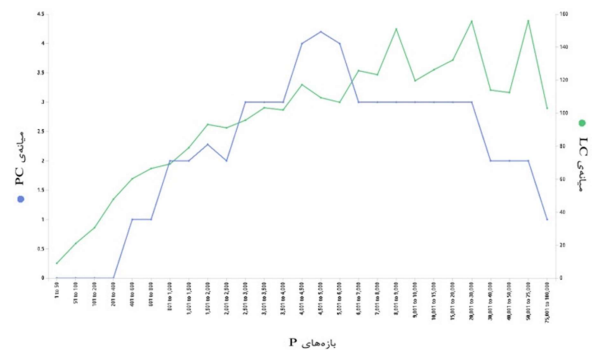
شکل ۷ توزیع پیام‌های تکراری در بازه‌های مختلف تعداد لینک

بین تعداد اعضای گروه‌ها و میانگین طول پیام‌های آن‌ها نیز رابطه وجود دارد. به‌طور کلی هر چه اعضای گروه بیشتر باشند، طول پیام‌های منتشر شده در آن گروه کوتاه‌تر است. به عبارت دیگر در گروه‌های شلوغ بیشتر از گروه‌های خلوت گفتگو صورت می‌گیرد. در گروه‌های خلوت چون گفتگو کمتر است، پیام‌های طولانی میانگین طول پیام را بالا می‌برند. شکل ۸ میانگین و میانه طول پیام را برای بازه‌های مختلف تعداد اعضا نشان می‌دهد. همان‌طور که انتظار می‌رفت، گروه‌های چت‌وگفتگو در بازه گروه‌های با پیام‌های کوتاه قرار گرفته و گروه‌های تخصصی در بازه پیام‌های طولانی‌تر هستند.



شکل ۸ میانگین و میانه طول پیام‌ها در بازه‌های تعداد اعضا

همان‌طور که در تحلیل رفتار گروه‌های تخصصی و چت‌وگفتگو بیان شد، گروه‌های تخصصی بیشتر از گروه‌های چت‌وگفتگو شماره تلفن منتشر می‌کنند. در شکل ۹ میانه تعداد لینک‌ها و تعداد شماره تلفن‌های منتشر شده در گروه‌ها بر اساس تعداد اعضای آن‌ها نمایش داده شده‌است. برخلاف روند تعداد لینک که با افزایش اعضای گروه بیشتر شده‌است، تعداد شماره تلفن‌ها تا گروه‌های با ۵۰۰۰ عضو افزایش داشته و سپس کاهش پیدا کرده است. با توجه به تحلیل رفتار مذکور، از این نمودار نتیجه می‌گیریم که تعداد شماره تلفن‌های منتشر شده در گروه می‌تواند معیار خوبی برای کیفیت گروه باشد. گروه‌هایی که خدمات تخصصی ارائه می‌دهند بیشتر شماره تلفن ارسال می‌کنند. در گروه‌های چت‌وگفتگو که اعضای بیشتری دارند و خدمات تخصصی ارائه نمی‌شود، شماره تلفن کمتری منتشر می‌شود. این موضوع برای تعداد لینک‌ها صادق نیست.



شکل ۹ میانه تعداد لینک‌ها و میانه تعداد شماره تلفن‌ها در بازه تعداد اعضا

۶- نتیجه‌گیری و پژوهش‌های آتی

در این پژوهش رفتار کاربران در سوپرگروه‌های تلگرامی بررسی شده‌است. نتیجه این بررسی، تعریف معیارهایی برای تعیین کیفیت سوپرگروه‌های تلگرامی است. ما به این نتیجه رسیدیم که سوپرگروه‌های باکیفیت شماره تلفن‌های بیشتری نسبت به بقیه منتشر می‌کنند. طول پیام‌ها در سوپرگروه‌های باکیفیت بیشتر است. پیام‌های فرورادی از کانال‌های متنوع‌تری هستند. کاراکترهای غیرعادی کمتری در نام این سوپرگروه‌ها مشاهده می‌شود. از همه مهم‌تر این‌که درگیری کاربران در بحث‌هایی که در سوپرگروه‌های باکیفیت می‌شود بیشتر از بقیه سوپرگروه‌ها است. نرخ پاسخ به پیام‌ها در سوپرگروه‌های باکیفیت نسبت به میانگین بیشتر است، ولی در چهارک اول نیست. سوپرگروه‌هایی که نرخ پاسخ آن‌ها در چهارک اول است، اکثراً بی‌کیفیت هستند.

برای تشخیص تفاوت‌های بین گروه‌های باکیفیت و کم‌کیفیت، دادگان کوچکی متشکل از دو موضوع (تخصصی و چت‌وگفتگو) ایجاد کردیم. در تلگرام برای موضوعات گسترده‌ای سوپرگروه وجود دارد.

معیارهایی که به‌دست آورده‌ایم می‌توانند به ما کمک کنند تا سوپرگروه‌ها را در موضوعات مختلف خوشه‌بندی کنیم. با توجه به این‌که ما متن همه پیام‌های سوپرگروه‌ها را ذخیره می‌کنیم، بازیابی کلماتی که بیشترین TF/IDF را در بین تمامی پیام‌ها دارند به‌آسانی مقدور است. با به‌دست آوردن این کلمات می‌توانیم خوشه‌بندی را هم بر روی معیارها و هم بر روی این کلمات انجام دهیم. با ترکیب نتایج می‌توانیم معیارها را دقیق‌تر بر اساس موضوع موشکافی کنیم.

مراجع

- [۱] مهدی بیگلرلو، بررسی رابطه فعالیت در تلگرام با فرهنگ سیاسی استادان دانشگاه‌های کلانشهرهای ایران، جستارهای سیاسی معاصر، دوره ۹، شماره ۲، صفحات ۱-۳۱، ۱۳۹۷.
- [۲] زهرا پیرجلیلی، سید محمد صادق مهدوی و مصطفی ازکیا، آسیب شناسی تأثیر شبکه اجتماعی تلگرام بر روابط خانوادگی زوجین (شهر تهران سال ۱۳۹۵)، فصلنامه بررسی مسائل اجتماعی ایران، دوره ۹، شماره ۱، صفحات ۶۹-۹۰، ۱۳۹۷.
- [۳] محمد عرب طاهری، علی وفاجو و عباس علیاری، تأثیر پیامدهای روانی شبکه اجتماعی تلگرام بر ارتکاب جرائم جوانان با رویکرد پیشگیری از جرم، فصلنامه پژوهش‌های انتظام اجتماعی، دوره ۹، شماره ۴، صفحات ۲۰۱-۲۲۰، ۱۳۹۶.
- [۴] سعید فرامرزیانی، شهناز هاشمی و علی اکبر فرهنگی، نقش رسانه های مجازی در تغییر ارزشهای اجتماعی با تأکید بر شبکه‌های اجتماعی تلگرام و فیسبوک، توسعه اجتماعی، دوره ۱۰، شماره ۴، صفحات ۱۲۳-۱۴۸، ۱۳۹۵.
- [۵] مهدیه رضائی و سید محمد شبیری، تأثیر کاربرد شبکه‌های اجتماعی بر ارتقای رفتار گردشگری حامی محیط زیست (مطالعه موردی: شبکه اجتماعی تلگرام)، برنامه‌ریزی و توسعه گردشگری، دوره ۶، شماره ۲۱، صفحات ۲۸-۵۳، ۱۳۹۶.
- [۶] حسین کرمانی، تلگرام و دیگر هیچ: تحلیل و تبیین عوامل مؤثر بر استفاده از تلگرام به عنوان پیام‌رسان مرجع در ایران، مطالعات رسانه‌ای، دوره ۱۳، شماره ۱، صفحات ۷-۲۰، ۱۳۹۷.
- [۷] فرهاد زیویار، مهدی شاهین، جعفر حسن‌خانی و رضا رحمتی، نقش شبکه‌های اجتماعی در هویت تمدنی کاربران ایرانی (مطالعه موردی فیسبوک و تلگرام)، مطالعات بین‌رشته‌ای رسانه و فرهنگ، دوره ۷، شماره ۱۴، صفحات ۱۵۰-۱۲۸، ۱۳۹۶.
- [۸] سید نصرت‌الله نعمتی‌فر و سید علی‌محمد رضوی، مهارت‌های شناختی تفکر انتقادی در کاربران شبکه‌های اجتماعی (مورد مطالعه شبکه اجتماعی تلگرام)، مطالعات رسانه‌ای، دوره ۱۳، شماره ۱، صفحات ۶۷-۸۵، ۱۳۹۷.
- [۹] سید محمد شبیری، نقش شبکه‌های اجتماعی بر فرهنگ‌سازی ارزشها و باورهای محیط‌زیستی معلمان: مطالعه موردی تلگرام، فصلنامه علمی-پژوهشی آموزش محیط‌زیست و توسعه پایدار، دوره ۶، شماره ۲، صفحات ۳۷-۴۳، ۱۳۹۶.
- [۱۰] مسیح‌السادات هدایت‌مفیدی، عطیه کامیابی‌گل و علی علی‌زاده، فضای مجازی و زبان فارسی: غیر معیارهای نوشتاری در شبکه اجتماعی تلگرام، مطالعات رسانه‌ای، دوره ۱۲، شماره ۳، صفحات ۶۵-۸۳، ۱۳۹۷.
- [۱۱] مسیح‌السادات هدایت‌مفیدی، عطیه کامیابی‌گل و علی علی‌زاده، آسیب‌شناسی زبانی و بلاغی زبان فارسی در شبکه ارتباطی تلگرام، توسعه اجتماعی، دوره ۱۰، شماره ۴، صفحات ۱۵۱-۱۶۸، ۱۳۹۶.
- [۱۲] ابوالفضل دانایی و الهام مؤمن، بررسی تأثیر بازاریابی ویروسی در شبکه‌های اجتماعی بر قصد خرید مصرف‌کنندگان (مورد مطالعه: کاربران شبکه اجتماعی تلگرام)، دو فصلنامه علمی-پژوهشی کوش‌های مدیریت بازرگانی، دوره ۹، شماره ۱۸، صفحات ۲۴۳-۲۶۷، ۱۳۹۶.
- [۱۳] حمیدرضا سعیدنیا و داوود قربانزاده، بررسی عوامل مؤثر بر نگرش و تبلیغات شفاهی مثبت در بین کاربران تلگرام، فصلنامه علمی-پژوهشی تحقیقات بازاریابی نوین، دوره ۷، شماره ۱، صفحات ۱۷۳-۲۰۲، ۱۳۹۶.
- [14] A. D. Nobari, N. Reshadatm and M. Neshati, Analysis of Telegram, An Instant Messaging Service, in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 2017.

زیر نویس‌ها

Cluster ^۸	Crawler ^۱
Chunk ^۹	Instance ^۲
Indexer ^{۱۰}	Pool ^۳
Lucene ^{۱۱}	Scheduler ^۴
Inverted Index ^{۱۲}	Fault Tolerant ^۵
Query ^{۱۳}	Seed ^۶
	Event ^۷