

به نام خدا



دانشگاه صنعتی شاهرود  
Shahrood University of Technology

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

عنوان تمرین : Face Book Anaysis

استاد: دکتر محسن رضوانی

نویسنده: مریم درویشیان

تاریخ: 1402/01/18

## آناتومی نمودار اجتماعی فیس بوک

ما ساختار نمودار اجتماعی کاربران فعال فیس بوک را مطالعه می کنیم، بزرگترین شبکه اجتماعی که تا کنون تحلیل شده است. ما ویژگی های متعددی از نمودار را محاسبه می کنیم که شامل تعداد کاربران و دوستی ها، توزیع درجه، طول مسیر، خوشه بندی و الگوهای ترکیبی می شود. نتایج ما حول محور سه نوع مشاهدات اصلی است. ابتدا، ساختار جهانی نمودار را مشخص می کنیم و مشخص می کنیم که شبکه اجتماعی تقریباً به طور کامل متصل است، با 99.91٪ افراد متعلق به یک جزء بزرگ متصل، و پدیده "شش درجه جدایی" را در مقیاس جهانی تأیید می کنیم. دوم، با مطالعه میانگین ضریب خوشه بندی محلی و هم تراز همسایگی های گراف، نشان می دهیم که در حالی که نمودار فیس بوک بطور کلی به وضوح پراکنده است، همسایه های نموداری کاربران دارای ساختار به طور شگفت آوری متراکم هستند. سوم، ما الگوهای دسته بندی موجود در نمودار را با مطالعه ویژگی های جمعیت شناختی و شبکه ای کاربران مشخص می کنیم.

ما درجه بندی واضحی را مشاهده می کنیم و میزان "دوستان شما دوستان بیشتری از شما دارند" را مشخص می کنیم. علاوه بر این، ما تأثیر قوی سن را بر ترجیحات دوستی و همچنین ساختار جامعه مدولار جهانی مبتنی بر ملیت مشاهده می کنیم. ما نتایج خود را با نتایج اجتماعی کوچکتر مقایسه می کنیم. شبکه ها و اکثراً، اما نه به طور کامل، توافق بر سر ویژگی های مشترک ساختاری شبکه پیدا می کنند.

## معرفی

ظهور سرویس های شبکه های اجتماعی آنلاین در دهه گذشته، نحوه مطالعه دانشمندان علوم اجتماعی بر ساختار روابط انسانی را متحول کرده است [1]. همانطور که افراد روابط اجتماعی خود را آنلاین می کنند، نقطه کانونی اینترنت از شبکه ای از اسناد به شبکه ای از افراد تبدیل می شود و ساختارهای اجتماعی نامرئی قبلی در مقیاسی فوق العاده و با جزئیات بی سابقه ای به تصویر کشیده می شوند. در این کار، ما ساختار بزرگترین شبکه اجتماعی آنلاین جهان، فیس بوک، را در تلاش برای پیشرفت وضعیت هنر در مطالعه تجربی شبکه های اجتماعی توصیف می کنیم. در ساده ترین شکل، یک شبکه اجتماعی شامل افراد به عنوان راس و یال ها به عنوان روابط بین رئوس است [2]. این نگاه انتزاعی به روابط انسانی، در عین اینکه قطعاً محدود است، برای آن بسیار مفید است.

توصیف روابط اجتماعی، با معیارهای ساختاری این انتزاع شبکه، کاربرد فعالی برای مطالعه همه چیز از قدرت چانه زنی [3] تا سلامت روانی [4] پیدا می کند. علاوه بر این، مشاهده شده است که شبکه های اجتماعی طیف گسترده ای از ویژگی های ساختاری یکپارچه را نشان می دهند، از جمله هموفیلی، خوشه بندی، اثر دنیای کوچک، توزیع های ناهمگون دوستان، و ساختار جامعه [5، 6].

تجزیه و تحلیل کمی از این روابط، مستلزم آن است که افراد به صراحت شبکه های اجتماعی خود را شرح دهند. از لحاظ تاریخی، مطالعات شبکه های اجتماعی به صدها نفر به عنوان داده های اجتماعی محدود می شد و روابط از طریق ابزارهای سخت و دشوار جمع آوری میشد. شبکه های اجتماعی آنلاین به ما امکان می دهند مقیاس و دقت چنین مطالعاتی را به طور چشمگیری افزایش دهیم، زیرا داده های جدید شبکه های اجتماعی، عمدتاً از منابع آنلاین، روابط اجتماعی ما را در مقیاس تقریباً جهانی ترسیم می کنند. مطالعات قبلی در مورد شبکه های اجتماعی آنلاین شامل تحقیق در توییتر، فلیکر، یاهو! 360، Myspace، Cyworld، LiveJournal [7-11] و Orkut.

ترند در این خط از تحقیقات اندازه گیری بازنمایی های بزرگتر و بزرگتر از شبکه های اجتماعی، از جمله شبکه های مشتق شده از ایمیل [12]، تلفن [13] و پیام های فوری [14] است. دو تحقیق اخیر روی رنرن [15] و پیام رسان MSN [14] به ترتیب شامل 42 میلیون و 180 میلیون نفر بودند. کامل بودن شبکه به ویژه در مطالعه شبکه های اجتماعی آنلاین اهمیت دارد زیرا بر خلاف تحقیقات سنتی علوم اجتماعی، اعضای شبکه های اجتماعی آنلاین به صورت تصادفی کنترل نمی شوند.

نمونه ها، و در عوض باید نمونه های مغرضانه در نظر گرفته شوند. در حالی که جمعیت شناسی این شبکه ها شروع به نزدیک شدن به جمعیت شناسی جمعیت جهانی به طور کلی کرده است [16]، دقیق ترین نمایش روابط اجتماعی ما تا حد امکان شامل افراد بیشتری خواهد بود. در این مقاله کل شبکه اجتماعی اعضای فعال فیس بوک را در می 2011 مشخص می کنیم، شبکه ای که در آن زمان از 721 میلیون کاربر فعال تشکیل شده بود. تا جایی که ما می دانیم، این بزرگترین شبکه اجتماعی است که تا کنون تحلیل شده است.

فیس بوک به طور طبیعی در گذشته توجه محققان را به خود جلب کرده است. برخی از این تحقیقات به درک زیرمجموعه های کوچک جمعیت فیس بوک، از جمله به طور خاص شبکه های اجتماعی دانشجویان دانشگاه اختصاص داده شده است [17-19]. مطالعات دیگر الگوهای ارتباطی و فعالیت را در میان بخش هایی از جمعیت کاربر تحلیل کرده اند [20،21]. سپس رشته دیگری از تحقیقات برخی از ویژگی های شبکه در مقیاس بزرگ نمودار فیس بوک را از طریق نمونه گیری و روش های دیگر برای جمع آوری داده های شبکه اندازه گیری کرده است [22،23]. قابل ذکر است، این روش ها راهی برای تمایز بین حساب های فعال و قدیمی ندارند. برخلاف این مطالعات، ما کل نمودار فیس بوک را به صورت ناشناس تجزیه و تحلیل می کنیم و بر مجموعه ای از حساب های کاربری فعال که به طور قابل اعتماد مربوط به افراد هستند تمرکز می کنیم. اگر کاربر فیس بوک در 28 روز گذشته از زمان اندازه گیری ما در می 2011 وارد سایت شده باشد و حداقل یک دوست فیس بوک داشته باشد، ما به عنوان یک عضو فعال شبکه اجتماعی تعریف می کنیم. توجه داریم که این تعریف دقیقاً تعریف معمولی فیس بوک از کاربر فعال نیست و بنابراین برخی از آمارهای ما کمی با آمار شرکت متفاوت است. طبق تعریف ما از فعال، جمعیت کاربران فعال فیس بوک در زمان اندازه گیری های ما حدود 721 میلیون نفر بود. برای مقایسه، تخمین زده می شود که جمعیت جهان حدود 6.9 میلیارد نفر در می 2011 این کشور است. به این معنی است که این شبکه تقریباً 10 درصد از جمعیت جهان را شامل می شود. محدود کردن مقایسه به افراد 13 ساله یا بیشتر و با دسترسی به اینترنت (مجموعه افراد واجد شرایط داشتن حساب های فیس بوک) این درصد را به طور قابل توجهی بیشتر می کند. در زمان اندازه گیری ما 68.7 میلیارد یال دوستی وجود داشت، بنابراین میانگین کاربر فیس بوک در مطالعه ما حدود 190 دوست فیس بوک داشت. تجزیه و تحلیل ما قبل از وجود "اشتراک" در فیس بوک بود، و در این کار ما فقط دوستی های متقابل فیس بوک را مطالعه می کنیم. ما همچنین زیرگراف 149 میلیون کاربر آمریکایی فیس بوک را تجزیه و تحلیل کردیم. با استفاده از تخمین های جمعیت از اداره سرشماری ایالات متحده برای سال 2011، تقریباً 260 میلیون نفر در ایالات متحده بالای سن 13 وجود دارد و بنابراین واجد شرایط ایجاد یک حساب کاربری فیس بوک است. بنابراین، در داخل ایالات متحده، شبکه اجتماعی فیس بوک بیش از نیمی از جمعیت واجد شرایط را شامل می شود. این زیرجمعیت 15.9 میلیارد لبه داشت، بنابراین متوسط کاربر ایالات متحده با حدود 214 کاربر دیگر ایالات متحده دوست بود. این مقدار میانگین بالاتر ممکن است نشان دهنده پذیرش عمیق تر فیس بوک در ایالات متحده از می 2011 باشد.

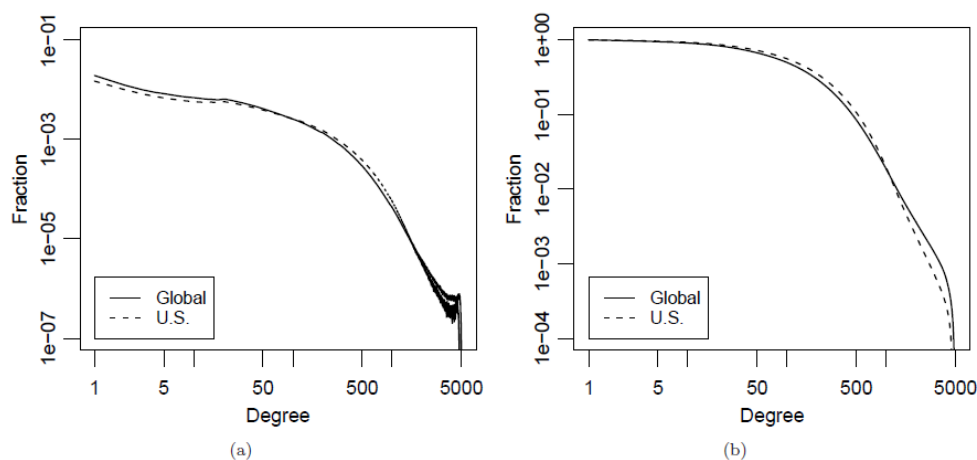
اهداف ما از توصیف شبکه اجتماعی فیسبوک دوگانه است. اول، هدف ما ارتقاء دانش جمعی از شبکه های اجتماعی و ارضای کنجکاوی گسترده در مورد روابط اجتماعی است. در فیس بوک تجسم یافته است. دوم، ما امیدواریم توسعه الگوریتم های نمودار و

ابزارهای تحلیل شبکه را به سمت نمایش واقعی این روابط متمرکز کنیم. برای رسیدن به این اهداف، ما در اینجا توضیح دقیقی از شبکه اجتماعی فیس بوک ارائه می دهیم.

### نتایج

در این بخش، ما طیف گسترده ای از معیارهای نمودار را در شبکه اجتماعی فیس بوک اعمال می کنیم. از آنجایی که شبکه واقعاً عظیم است، ما از منابع محاسباتی گسترده ای برای انجام این اندازه گیری ها استفاده می کنیم. با این حال، تمرکز این مقاله بر نتایج این اندازه گیری ها است، و بنابراین ما بحث های تکنیکی خود را به روش ها واگذار می کنیم.

3



شکل 1. توزیع درجه

جمعیت ایالات متحده از کاربران فیس بوک. (ب) تابع توزیع تجمعی مکمل (CCDF).

نمودار فیس بوک

توزیع درجه یک کمیت اساسی است که به طور مکرر در مطالعات تجربی شبکه ها اندازه گیری شده است، توزیع درجه  $p_k$  است. درجه  $k$  یک فرد، تعداد دوستانی است که فرد دارد و  $p_k$  کسری از افراد در شبکه است که دقیقاً  $k$  دوست دارند. ما توزیع درجه کاربران فعال فیس بوک را در کل جمعیت جهان و همچنین در زیر جمعیت کاربران آمریکایی محاسبه کردیم. توزیع درجه جهانی و ایالات متحده در شکل 1 نشان داده شده است که در مقیاس  $\log$ - $\log$  نمایش داده شده است.

از آنجایی که توزیع برای ایالات متحده کاملاً مشابه توزیع کل جمعیت است، ما توجه خود را بر توزیع درجه جهانی متمرکز می کنیم. توزیع تقریباً یکنواخت کاهش می یابد، به جز یک ناهنجاری کوچک در نزدیکی 20 دوست. این

پیچیدگی به دلیل نیروهایی است که در محصول فیس بوک وجود دارد که به ویژه افراد با تعداد کم دوستان را تشویق می کند تا دوستان بیشتری پیدا کنند تا زمانی که به 20 دوست برسند. توزیع یک بریدگی واضح در 5000 دوست محدودیتی که فیس بوک بر تعداد دوستان در زمان اندازه گیری ما اعمال می کند را نشان می دهد. توجه داشته باشید که از آنجایی که 5000 به هیچ وجه به تعداد کاربران فیس بوک نزدیک نیست، هر کاربر به وضوح با بخشی از جمعیت فیس بوک دوست است. با بازتاب بیشتر شبکه های اجتماعی مشاهده شده، روابط اجتماعی ما پراکنده است.

در واقع، اکثر افراد تعداد متوسطی از دوست کمتر از 200 را در فیس بوک دارند، در حالی که جمعیت بسیار کمتری، صدها یا حتی هزاران دوست دارند، دوست متوسط حساب می شود. کاربران جهانی در مطالعه ما 99 نفر بودند. جمعیت کوچکی از کاربران با درجات غیرعادی بالا که گاهی اوقات در ادبیات شبکه ها هاب نامیده می شود، درجاتی بسیار بزرگتر از میانگین یا متوسط کاربر فیس بوک دارند. توزیع به وضوح دارای انحراف راست با واریانس بالا است، اما قابل توجه است که انحنای قابل توجهی در توزیع در مقیاس  $\log\text{-}\log$  نشان داده شده است. این انحنا تا حدودی شگفتانگیز است، زیرا اندازه گیری های تجربی شبکه ها اغلب ادعا می کنند که توزیع های درجه به اصطلاح قدرت را دنبال می کنند.

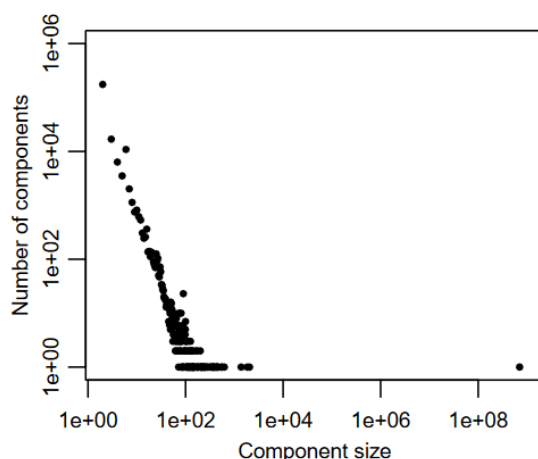
در واقع، بیشتر افراد تعداد متوسطی از دوستان در فیس بوک دارند، کمتر از 200 نفر، در حالی که جمعیت بسیار کمتری صدها یا حتی هزاران دوست دارند. دوست متوسط برای کاربران جهانی در مطالعه ما 99 نفر بودند. جمعیت کمی از کاربران با درجه های غیر طبیعی بالا، که گاهی اوقات در ادبیات شبکه ها به نام Hubs نامیده می شوند، دارای درجه بسیار بیشتری از کاربر متوسط یا متوسط فیس بوک هستند. توزیع به وضوح با واریانس بالایی مناسب است، اما قابل توجه است که انحنای قابل توجهی در توزیع در مقیاس ورود به سیستم وجود دارد. این انحنای تا حدودی تعجب آور است، زیرا اندازه گیری های تجربی شبکه ها اغلب ادعا کرده اند که توزیع درجه برای پیروی از به اصطلاح نیروگاه ها، که از نظر ریاضی توسط  $PK \propto K^{-\alpha}$  برای برخی  $\alpha > 0$  نشان داده شده است [24، 25]. قوانین قدرت خطوط مستقیم در یک نقشه ورود به سیستم هستند و به وضوح توزیع مشاهده شده مستقیم نیست. نتیجه می گیریم، مانند Ref. [23]، این مدل های سخت قانون قدرت برای توزیع درجه فیس بوک نامناسب هستند. با این وجود هدف ما این نیست که مشخص کنیم کدام فرم پارامتری توزیع را بهترین مدل می کند. نتایج مربوطه یکنواخت است و انحنای توزیع درجه، درجه کاربران معمولی، واریانس بزرگ در درجه ها و کمبود شبکه. با این حال، کمبود شبکه دلالت بر این ندارد که کاربران در شبکه فیس بوک از یکدیگر دور هستند. در حالی که بیشتر جفت کاربران به طور مستقیم به یکدیگر متصل نیستند، اما تقریباً همه جفت کاربران از طریق مسیرهای طول طولانی تر متصل می شوند. در بخش بعدی فاصله بین کاربران در طول مسیر نمودار اجتماعی را اندازه گیری می کنیم.

هنگام مطالعه ساختار شبکه، توزیع فاصله بین راس ها یک خاصیت ماکروسکوپی با علاقه اساسی است. در اینجا ما توابع همسایگی و مسافت متوسط جفت شبکه های فیس بوک و ایالات متحده را مشخص می کنیم. به طور رسمی، تابع همسایگی  $H(N)$  از یک نمودار تعداد جفت های راس  $(U, V)$  را به گونه ای توصیف می کند که  $U$  از  $V$  در طول مسیر در شبکه با لبه های  $H$  یا کمتر قابل دسترسی باشد. با توجه به تابع همسایگی، قطر یک نمودار به سادگی حداکثر فاصله بین هر جفت رئوس موجود در نمودار است. قطر یک اندازه گیری افراطی است.

تابع همسایگی ، که اندازه گیری صدک جفت راس در فاصله مشخصی است. قطر دقیق را می توان با حضور یک مسیر سخت در برخی از مناطق محیطی نمودار به شدت تغییر داد ، در حالی که تصور می شود که تابع همسایگی و میانگین آن ، فاصله های "معمولی" بین جفت راس را به شدت میگیرد. مانند بسیاری از نمودارهای دیگر ، نمودار Facebook بین همه جفت راس ها مسیر ندارد. این مانع از توصیف شبکه با استفاده از تابع همسایگی نمی شود. همانطور که در بخش بعدی خواهیم دید ، اکثریت قریب به اتفاق شبکه از یک مؤلفه بزرگ متصل تشکیل شده است و بنابراین تابع همسایگی نماینده اکثریت قریب به اتفاق جفت های رئوس است.

شکل 2 عملکرد همسایگی محاسبه شده برای هر دو نمودار همه کاربران فیس بوک و همچنین نمودار کاربران فیس بوک ایالات متحده را نشان می دهد ، از ماه مه 2011 ، با استفاده از الگوریتم های HyperAnf که اخیراً توسعه یافته است [48]. ما می دانیم که میانگین فاصله بین جفت کاربران برای کاربران فیس بوک 4.7 و برای کاربران ایالات متحده 4.3 بود. طول مسیر کوتاه بین افراد ، به اصطلاح "شش درجه جدایی" که توسط آزمایش های استنلی میلگرام در مورد بررسی شبکه اجتماعی ایالات متحده وجود دارد [26] ، در اینجا در فیس بوک در مقیاس جهانی مشاهده می شود. همانطور که در شکل 2 نشان داده شده است ، 92٪ از کل جفت کاربران فیس بوک در پنج درجه جدایی و 99.6٪ در شش درجه بودند. با توجه به شبکه اجتماعی تنها کاربران ایالات متحده ، 96٪ در پنج درجه و 99.7٪ در شش درجه بودند. برای جزئیات فنی این محاسبات ، خواننده علاقه مند را به یک مقاله جداگانه در مورد فشرده سازی و عبور از نمودار فیس بوک ارجاع می دهیم [27].

5

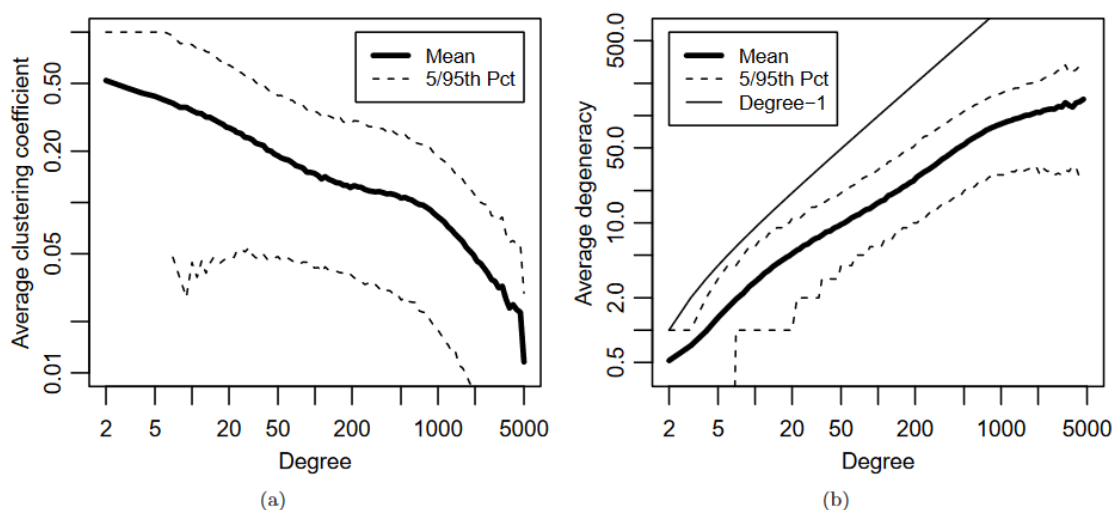


**Figure 3. Component size distribution.** The fraction of components with a given component size on a log-log scale. Most vertices (99.91%) are in the largest component.

به منظور نشان دادن این که این نتایج طول مسیر نماینده کل شبکه فیس بوک است ، اکنون ساختار مؤلفه نمودار را بررسی می کنیم. اندازه مؤلفه نتیجه گیری ما از بخش قبلی مبنی بر اینکه شبکه میانگین بسیار کوتاه دارد. طول مسیر به وجود چنین مسیرهایی بین بیشتر جفت های رئوس متکی است ، واقعیتی که ما در این بخش تأیید خواهیم

کرد. ما این کار را با یافتن مؤلفه های متصل شبکه اجتماعی ، انجام می دهیم جایی که  $a$  مؤلفه متصل مجموعه ای از افراد است که برای آنها هر جفت از افراد حداقل با یک مسیر از طریق شبکه متصل می شوند. محاسبات عملکرد محله ما فقط فاصله بین جفت کاربران در مؤلفه های متصل را محاسبه می کند زیرا این تنها کاربرانی هستند که در واقع از طریق مسیرها متصل هستند. برای اینکه نتایج حاصل از بخش قبلی به عنوان توصیف قطر تعبیر شود ، ما نیاز داریم که بیشتر ، اگر نه همه ، شبکه در یک مؤلفه بزرگ متصل باشد.

در شکل 3 ، ما توزیع اندازه مؤلفه ها را در مقیاس های ورود به سیستم نشان می دهیم ، که دقیقاً با استفاده از الگوریتم شرح داده شده در روش ها یافت می شود. در حالی که بسیاری از مؤلفه های متصل وجود دارد ، بیشتر این ترکیبات نانت ها بسیار کوچک هستند. دومین مؤلفه متصل فقط بیش از 2000 نفر را دارد ، در حالی که بزرگترین مؤلفه متصل ، تمام راه در سمت راست شکل ، شامل 99.91 درصد درصد شبکه است. این مؤلفه شامل اکثریت قریب به اتفاق کاربران فعال فیس بوک با حداقل یک دوست است. بنابراین نه تنها طول مسیر متوسط بین افراد کوتاه است ، بلکه این ارتباطات اجتماعی بین همه افراد در فیس بوک وجود دارد. طول مسیر و ساختار مؤلفه شبکه ، نمایی از شبکه را در مقیاس ماکروسکوپی به ما می دهد و اکنون با بررسی خصوصیات محلی بیشتر شبکه ، تحقیقات خود را ادامه می دهیم. ضریب خوشه بندی و هم تراز.



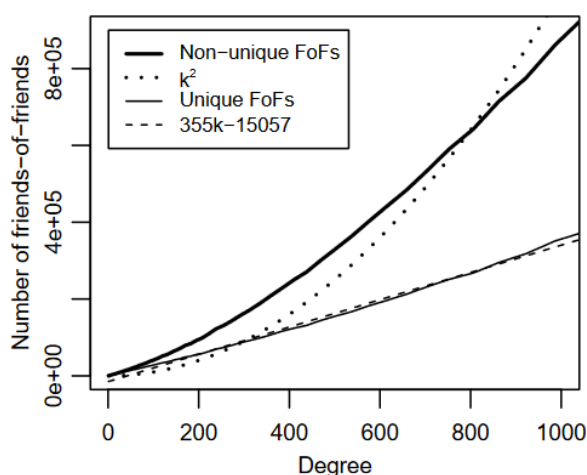
**Figure 4. Local clustering coefficient and degeneracy.** The clustering coefficient (a) and the degeneracy (b) as a function of degree on log-log scales. The means are shown as solid lines and the dashed lines display the 5/95th percentiles. An upper-bound on the degeneracy given by degree minus 1 is also shown as the thin solid line for comparison.

در ابتدا با محاسبه توزیع درجه ، تعداد دوستان هر کاربر را مشخص کردیم و اکنون تجزیه و تحلیل دقیق تری از همسایگی های گرافیکی اجتماعی کاربران انجام می دهیم. نمودار همسایگی برای کاربر  $i$  ، که گاهی اوقات نمودار

Ego یا 1 توپ نامیده می شود ، راس زیرگراف ناشی از کاربرانی که با کاربر A است و دوستی بین این کاربران دوست هستند. کاربر من در همسایگی خودشان گنجانده نشده است.

ما ابتدا ضریب متوسط خوشه بندی محلی را برای کاربران به عنوان تابعی از درجه محاسبه کردیم ، که برای راس درجه  $k$  درصد دوستی احتمالی بین دوستان  $K$  خود را اندازه گیری می کند (حداکثر  $(K - 1)/2$ ) در نمودار همسایگی خود وجود دارد. این نتیجه در شکل A4 نشان داده شده است ، توجه می کنیم محورها به صورت ورود به سیستم هستند.

ما می بینیم که ضریب خوشه بندی همسایگی صرف نظر از درجه ، در مقایسه با درصد دوستی های احتمالی در شبکه به طور کلی و مهمتر از همه ، در مقایسه با MEA- عملکرد سایر شبکه های اجتماعی آنلاین بررسی شده. به عنوان مثال ، برای کاربران با 100 دوست ، میانگین ضریب خوشه بندی محلی 0.14 است که نشان می دهد برای یک کاربر متوسط ، 14٪ از کل جفت های دوستشان خودشان دوست هستند. این تقریباً پنج برابر بیشتر از ضریب خوشه بندی موجود در یک مطالعه 2008 است که با تجزیه و تحلیل نمودار مکاتبات پیام رسان MSN ، برای همان اندازه محله [14]. در همین حال ، تجزیه و تحلیل ما همچنین نشان می دهد که ضریب خوشه بندی به صورت یکنواخت با درجه ، مطابق با مطالعه قبلی پیام رسان MSN و سایر مطالعات کاهش می یابد. به طور خاص ، ضریب خوشه بندی.



**Figure 5. Friends-of-friends.** The average number of unique and non-unique friends-of-friends as described in the text as a function of degree. Degree squared and a linear fit of the unique friends-of-friends are shown for comparison as the dotted and dashed lines.

به سرعت برای کاربران با نزدیک به 5000 دوست خلاصه می شود ، و این نشان می دهد که این کاربران به احتمال زیاد از فیس بوک برای اهداف اجتماعی کمتری استفاده می کنند و کاربران را به طور غیرقانونی دوست دارند.

با مشاهده چنین ضرایب بزرگ خوشه بندی در همسایگیهای محلی ، ما تصمیم گرفتیم با اندازه گیری هم ترازای آنها ، کمبود نمودارهای همسایگی را بیشتر مطالعه کنیم. به طور رسمی ، هم ترازای یک نمودار  $G$  بدون هدایت بزرگترین



K است که G برای یک هسته K غیر خالی است [28]. در همین حال ، هسته k از نمودار g حداکثر زیرگراف G است که در آن تمام راس ها دارای درجه حداقل k هستند ، یا به طور معادل آن ، زیرگراف G تشکیل شده با حذف تکراری تمام راس های درجه کمتر از k تا همگرایی. حداکثر k هسته یک نمودار G شباهت مفهومی به حداکثر  $k-1$  clique را تحمل می کند ، اما توجه به این نکته حائز اهمیت است که یک هسته k لزوماً  $k-1$  Clique نیست ، مگر اینکه K-Core دقیقاً حاوی  $k-1$  راس. K-Core با این حال نشانه ای به راحتی محاسبه و قوی از چگونگی وجود یک جامعه محکم در یک نمودار معین ارائه می دهد.

ما میانگین هم ترازی را به عنوان تابعی از درجه کاربر در شکل B4 گزارش می کنیم ، که دوباره در مقیاس ورود به سیستم ترسیم شده است. در نمودارهای همسایگی کاربران ، می فهمیم که میانگین هم ترازی تابع درجه کاربر در حال افزایش است. این باید با انتظارات ما سازگار باشد: هرچه دوستان بیشتری داشته باشید ، عضو یک جامعه محکم و محکم هستید که در آن قرار دارید. با این حال تعجب آور است که در واقع این همسایگی ها چقدر متراکم هستند: برای یک کاربر با 100 دوست ، میانگین هم ترازی همسایگی آنها 15 است. علاوه بر این ، برای کاربران با 500 دوست ، میانگین هم ترازی آنها 53 است ، به این معنی که حداقل 54 نفر دوستانی که همه 53 نفر از دوستان دیگر خود را می شناسند. در مقابل ، Eppstein و Strash اخیراً هم ترازی چندین نمودار ، چه اجتماعی و غیر اجتماعی را مورد بررسی قرار دادند و دریافتند که در کل نمودارها (که فقط همسایگی ها را بررسی نمی کنند) ، هم ترازی بسیار متوسط تر بودند [29]. برای نمودار 36,692 ورتکس ارتباطات ایمیل Enron ، هم ترازی تنها 43 بود.

هم ترازی ها به سادگی با در نظر گرفتن همسایگی یک کاربر متوسط با 500 دوست است. این نشان می دهد که اگرچه نمودار فیس بوک در کل پراکنده است، وقتی کاربران تعداد دوستان قابل توجهی را جمع می کنند، دوستی های آنها به دور از تبعیض نیست، و در عوض حول هسته های متراکم قابل توجهی متمرکز می شوند. اکنون با بررسی دوستان دوستان تک کاربران، همسایگی یک راس را تا فواصل دورتر در نظر می گیریم. دوستان دوستان ، یکی از ویژگی های مهم نمودارها که در طراحی الگوریتم ها باید در نظر گرفته شود.

تعداد راس هایی که در دو هاپ یک راس اولیه قرار دارند. این خاصیت تعیین می کند که الگوریتم های نمودار Traversal ، مانند جستجوی وسعت اول ، امکان پذیر هستند. در شکل 5 ، ما میانگین تعداد دوستان منحصر به فرد و غیر متحد دوستان را به عنوان تابعی از درجه محاسبه کردیم. تعداد دوستان غیرقانونی دوستان با تعداد مسیرهای طول دو با طول که از یک راس اولیه شروع می شوند و به آن راس باز نمی گردند مطابقت دارد. تعداد دوستان منحصر به فرد دوستان با تعداد راس های منحصر به فرد قابل دسترسی در انتهای یک مسیر طول دو مطابقت دارد. یک رویکرد ساده لوحانه برای شمارش دوستان دوستانه فرض می کند که یک کاربر با دوستان K رویکرد ساده لوحانه برای شمارش دوستان دوستان تقریباً دارد. فرض می کند که: کاربری با k دوست تقریباً  $k^2$  دوستان غیر منحصر به فرد دارد، با این فرض که دوستانش تقریباً همان دوستان را دارند.

این را می توان تخمین ساده لوحانه سخاوتمندانه ای از تعداد دوستان منحصر به فرد تلقی کرد، سخاوتمندانه زیرا در بالا دیدیم که بخش قابل توجهی از دوستان دوستان شما دوستان شما هستند. که در واقعیت، تعداد دوستان غیر منحصر به فرد دوستان فقط نسبتاً سریعتر از خطی رشد می کند، و تعداد دوستان منحصر به فرد دوستان بسیار نزدیک به خطی است، با یک تناسب خطی که شیب 355 دوست دوستان منحصر به فرد را در هر تولید می کند. در حالی که

نرخ رشد ممکن است کندتر از حد انتظار باشد، دوست اضافی تا زمانی که یک کاربر بیش از 800 دوست داشته باشد، مهم است از این شکل مشاهده کنید که مقادیر مطلق به طور غیرمنتظره ای بزرگ هستند: کاربری با 100 دوست، 27,500 دوست دوستان منحصر به فرد و 40,300 دوست دوستان غیر منحصر به فرد دارد. این به طور قابل توجهی بیشتر از  $99 * 100 = 9,900$  دوست غیر منحصر به فرد است که اگر دوستان ما تقریباً به اندازه ما دوست داشتند، انتظار داشتیم. این زیاده روی به اصلی مربوط می شود که در زیر به تفصیل در مورد آن بحث خواهیم کرد، جایی که نشان می دهیم «دوستان شما دوستان بیشتری نسبت به شما دارند»، یک نتیجه ثابت از مطالعات قبلی شبکه های اجتماعی است [30]. همبستگی های درجه تعداد دوستی ها در همسایگی شبکه همسایگی شما به تعداد دوستان، درجه و درجه دوستان شما بستگی دارد. در بسیاری از شبکه های اجتماعی آنلاین و آفلاین مورد مطالعه بوده است.

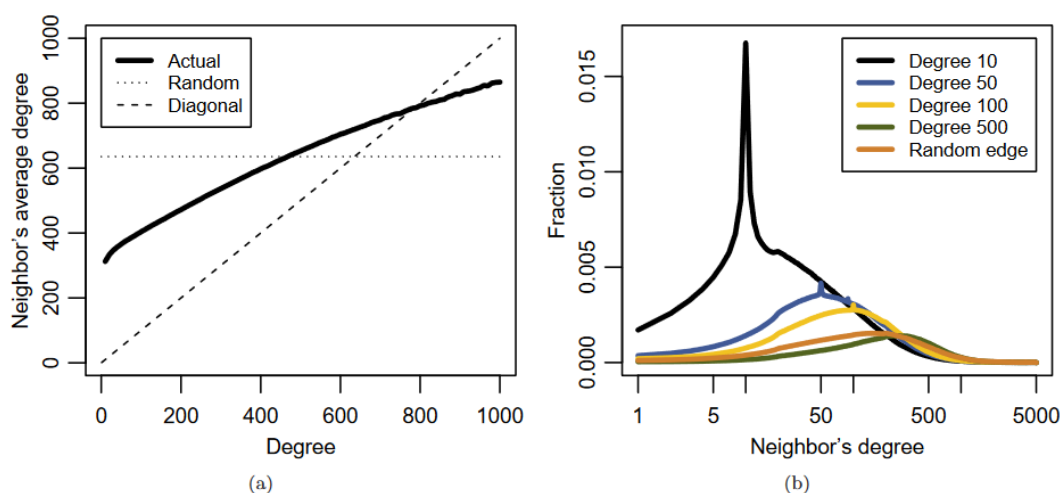
متوجه شده اید که درجه همسایه شما با درجه تحصیلی شما مرتبط است: وقتی درجه تحصیلی شما بزرگ است، این درجه بزرگ است و زمانی که درجه شما کوچک است، به اصطلاح به درجه طبقه بندی می گویند. ما می توانیم این همبستگی های درجه را با محاسبه ضریب همبستگی پیرسون  $r$  بین درجه های انتهای یک یال، کمی کنیم [31,32]. برای شبکه فیس بوک،  $r = 0.226$ ، نشان دهنده همبستگی مثبت با بزرگی مشابه با سایر نمودارهای اجتماعی است. این مقدار با مطالعات قبلی شبکه های کوچکتر از جمله همکاری دانشگاهی و همکاری بازیگران مطابقت دارد، جایی که مقادیر  $r$  از 0.120 تا 0.363 است [31]. معیار دقیق تر دیگر  $\langle k \rangle$ ، میانگین تعداد دوستان برای همسایه یک فرد با  $k$  دوست [33]، به عنوان خط ثابت در شکل 6a نشان داده شده است. (ما از علامت  $\langle x \rangle$  برای نشان دادن میانگین کمیت  $x$  استفاده می کنیم). تعداد مورد انتظار دوستان در انتهای یک یال به طور تصادفی انتخاب شده،  $\langle k \rangle / \langle k^2 \rangle = 635$  است. به صورت یک خط نقطه چین افقی نشان داده شده است. برخلاف این مقدار ثابت که انتظار ما در صورت عدم وجود همبستگی درجه است - خط ثابت از نزدیک به 300 برای افراد درجه پایین به نزدیک به 820 افزایش می یابد.

افراد با هزاران دوست که مجموعه مثبت شبکه را تایید می کنند. (اندازه گیری ها بعد از 1000 پر سر و صدا می شوند و برای وضوح، رقم را در این نقطه قطع می کنیم). مقایسه خط ثابت با خط چین مورب نشان می دهد که تا زمانی که شما نزدیک به 700 دوست نداشته باشید، همسایه (متوسط) شما دوستان بیشتری از شما دارد. این پدیده به طور مفصل مورد بحث قرار گرفته است. فلد [30] و فیس بوک این اثر را در مقیاس بزرگ نشان می دهد. این واقعیت که همسایگان متوسط ما دوستان بسیار بیشتری دارند نیز توضیح می دهد که چرا تخمین های ساده دوست ما در بخش قبل خیلی پایین مشاهده فلد مبنی بر اینکه «دوستان شما دوستان بیشتری نسبت به شما دارند» یک پارادوکس روانشناختی مهم است که برای دوستی هم به کار می رود. هنگامی که افراد خود را با دوستان خود مقایسه می کنند، از نظر مفهومی مناسب تر است که این مقایسه را نسبت به میانه دوستان خود قاب بندی کنیم، و این سؤال را به عنوان موضوع پرسیدن "رتبه طبقاتی" در بین همسالان خود روانشناسی کنیم [34]. بنابراین یافته های ما در رابطه با میانه شاید مهم تر باشد: مشاهده می کنیم که 83.6 درصد از کاربران دوستان کمتری نسبت به تعداد دوستان متوسط دوستان خود دارند. همه این افراد تجربه می کنند که بیش از نیمی از دوستانشان بیشتر از آنها دوستان دارند.

برای کامل تر شدن، ما همچنین یادآور می‌شویم که 92.7٪ از کاربران دوستان کمتری نسبت به میانگین تعداد دوستان خود دارند. با این حال، ما می‌توانیم بیش از اندازه‌گیری این آمار ساده و مشروط را مشخص کنیم.

احتمال  $p(k'|k)$  که دوستی که به طور تصادفی انتخاب شده از فردی با درجه  $k$  درجه  $k'$  دارد [35]. ما این را برای مقادیر با فاصله مساوی از  $k$ ، همه مضرب های ده، محاسبه کردیم و توزیع را برای چند عدد نشان دادیم. مقادیر مثال  $k$  در شکل 6b، همراه با توزیع در صورت عدم وجود همبستگی درجه، به عنوان مثال توزیع درجاتی که با دنبال کردن یک یال به طور تصادفی انتخاب شده است.

ابتدا توجه داشته باشید که محور افقی لگ است در حالی که محور عمودی خطی است. مطابق با شکل 6a، میانگین این توزیع‌ها به وضوح کمتر از میانگین توزیع نارنجی است که نشان‌دهنده ادامه دادن یک یال تصادفی در شبکه است، به جز خط سبز که نشان دهنده  $k = 500$  است.

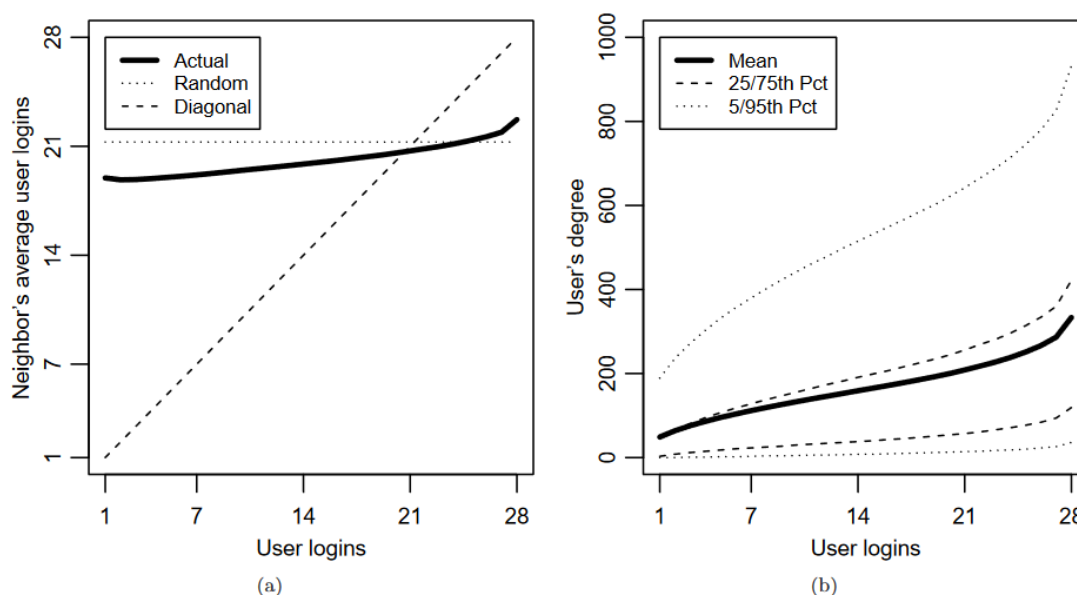


**Figure 6. Degree correlations.** (a) The average neighbor degree of an individual with degree  $k$  is the solid line. The horizontal dashed line shows the expected value if there were no degree correlations in the network  $\langle k^2 \rangle / \langle k \rangle$ , and the diagonal is shown as a dashed line. (b) The conditional probability  $p(k'|k)$  that a randomly chosen neighbor of an individual with degree  $k$  has degree  $k'$ . The solid lines, on the linear-log scale, show the measured values for four distinct degrees  $k$  shown in the caption. The orange line shows the expected distribution,  $\frac{k' p_{k'}}{\langle k \rangle}$ , if the degrees were uncorrelated.

دوباره، توزیع‌ها باز هم، با افزایش  $k$ ، توزیع‌ها به سمت راست تغییر می‌کنند که درجه طبقه‌بندی را نشان می‌دهد. علاوه بر این، با حذف هر گونه رفتار غیرهموار عجیب و غریب بین مقادیر نمونه‌گیری شده  $k$ ، میانه  $p(k'|k)$  تا بین 390 تا 400 دوست بزرگتر از  $k$  است، که تأیید می‌کند که رفتار میانگین در شکل 6a همراه کننده نیست مشاهدات دیگر از شکل، و داده‌های دیگر مقادیر  $k$  نشان داده نشده است، این است که درجه مودال دوستان دقیقاً برابر با  $k$  است تا حدود  $k = 120$ . بنابراین در حالی که دوستان شما احتمالاً دوستان بیشتری از شما دارند، به احتمال زیاد

تعداد دوستان همسایه شما با درجه تحصیلی شما برای کاربران درجه پایین تا متوسط یکسان است. همبستگی تعامل سایت. علاوه بر همبستگی درجه، همبستگی بین صفات افراد و ساختار شبکه را نیز بررسی کردیم [36]. ما اکنون محاسبات همبستگی خود را با استفاده از تعداد روزهایی که کاربران در طول پنجره 28 روزه مطالعه وارد شده اند، به جای درجه، که در شکل a7 مشاهده می شود، تکرار می کنیم. مجدداً، مقدار متوسط را در انتهای یک یال به طور تصادفی انتخاب شده و خط مورب را برای مقایسه ارائه می دهیم. برخلاف حالت درجه، در اینجا ابهام در تعریف همسایه تصادفی و از این رو میانگین تعداد ورود همسایه ها وجود دارد. تعریف ما از همسایه تصادفی رئوس با صفت  $X$  این است که ابتدا یک راس با صفت  $X$  را متناسب با درجه آنها انتخاب کنیم و سپس یک یال متصل به آن راس را به طور یکنواخت به صورت تصادفی انتخاب کنیم. به عبارت دیگر به هر یال متصل به رئوس با صفت  $X$  وزن مساوی می دهیم. بنابراین به رئوسی که به 5 راس با صفت  $X$  متصل است، 5 برابر وزن میانگین داده می شود.

یک راس که به یک راس منفرد با ویژگی  $X$  متصل می شود. مانند درجه، تعامل سایت مجاور شما با تعامل سایت شما در ارتباط است، اما میانگین تعداد ورود همسایه ها با انتظار تصادفی افقی بهتر از آنچه در مورد درجه دیده شد نشان داده می شود.

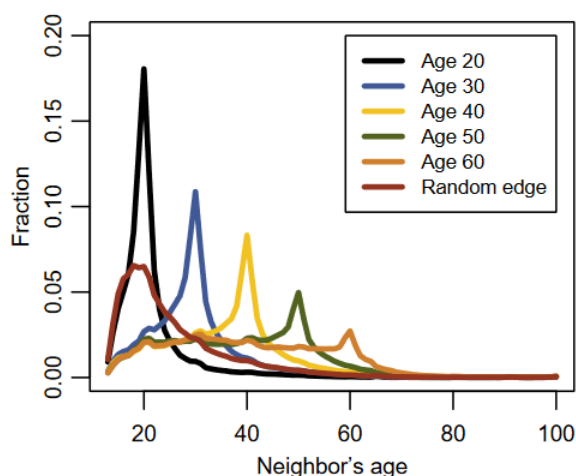


**Figure 7. Login correlations.** (a) Neighbor's logins versus user's logins to Facebook over a period of 28 days. The solid line shows the actual mean values and the horizontal line shows the average login value found by following a randomly chosen edge. The dashed line shows the diagonal. (b) A user's degree versus the number of days a user logged into Facebook in the 28 day period. The solid line shows the mean user degree, the dashed lines the 25/75 percentiles, and the dotted lines the 5/95th percentiles.

با این حال، مشاهده جالبتر این است که مقدار ثابت به مراتب بزرگتر از مقدار مورب روی بیشتر محدوده از ورود 0 تا 20 بار در 28 روز گذشته است. بنابراین با همان استدلال برای مورد درجه، تا زمانی که شما در حدود 70 درصد از روزهای یک ماه وارد سیستم شوید، دوستان شما بیشتر از شما وارد فیس بوک می شوند. ما می توانیم این پدیده را با

بررسی همبستگی بین درجه تحصیلی یک فرد و لاگین کردن به فیسبوک متوجه شویم. یک کاربر فیسبوک محتوا را از طریق بهروزرسانی وضعیت، پیوندها، ویدیوها، عکس‌ها و غیره به و از دوستان خود در شبکه اجتماعی ارائه و دریافت می‌کند و از این رو ممکن است در صورت داشتن دوستان بیشتر، انگیزه بیشتری برای ورود به سیستم داشته باشد. چنین همبستگی مثبتی بین درجه و ورود به سیستم وجود دارد، و ما آن را در شکل b7 نشان می‌دهیم. کاربری که به طور کلی وارد سیستم می‌شود، دوستان بیشتری در فیس بوک دارد و بالعکس. بنابراین از آنجایی که دوستان شما دوستان بیشتری نسبت به شما دارند، آنها نیز بیشتر از شما به الگوهای ترکیبی دیگر فیس بوک وارد می‌شوند. علاوه بر ورود به فیس بوک، ویژگی‌های کاربر دیگری نیز وجود دارد که می‌توان آنها را با ساختار شبکه مقایسه کرد. ما روی سه مقدار دیگر از این قبیل تمرکز می‌کنیم که اساساً پوشش کاملی برای کاربران فیس بوک دارند. سن، جنس، و کشور مبدأ [36] و الگوهای اختلاط [32] آنها. ما با در نظر گرفتن الگوهای دوستی در بین افراد با سنین مختلف شروع می‌کنیم و احتمال شرطی  $p(t'|t)$  را برای انتخاب یک همسایه تصادفی از افراد با سن  $t$  که سن  $t'$  دارند محاسبه می‌کنیم.

باز هم، همسایه تصادفی به این معنی است که به هر یال متصل به یک راس با سن  $t$  احتمال مساوی برای دنبال کردن داده می‌شود. ما این تابع را برای طیف وسیعی از مقادیر  $t$  در شکل 8 نمایش می‌دهیم. همانطور که ممکن است ساده لوحانه انتظار می‌رود و در عوض حدود مقدار حداکثر  $t' = t$  نامتقارن هستند. جای تعجب نیست که یک همسایه تصادفی به احتمال زیاد همسن شماست. بدیهی است که احتمال دوستی با افراد مسن‌تر به سرعت و تقریباً به‌طور تصاعدی از حالت خارج می‌شود. توزیع‌ها نیز در زیر حالت سقوط می‌کنند، توزیع‌ها نیز کاهش می‌یابند، اما سپس به مقداری نزدیک می‌شوند که تقریباً مستقل از سن کاربر  $t$  است (برای مثال خطوط آبی، زرد و سبز را ببینید). و از شکل متوجه می‌شویم که با افزایش  $t$  واریانس در توزیع افزایش می‌یابد. به طور کلی، افراد جوان بیشتر دوستان خود را در محدوده سنی کوچکی دارند در حالی که افراد مسن‌تر دارای طیف بسیار گسترده‌تری هستند.



**Figure 8.** The distribution  $p(t'|t)$  of ages  $t'$  for the neighbors of users with age  $t$ . The solid lines show the measured distributions against the age  $t$  described in the legend, and the red line shows the distribution of ages found by following a randomly chosen edge in the network.

هیچ یک از این رفتارها در مقایسه با توزیع سنی در انتهای یک لبه تصادفی انتخاب شده، خط قرمز، که در مرکز آن 20 است، مشهود نیست. بنابراین، اگرچه واضح است که سن برای روابط اجتماعی ما اهمیت دارد، شبکه اجتماعی فیس بوک نشان می دهد. الگوهای نامتقارن غیر پیش پا افتاده، سازگار در سنین کاربر  $t$ . با تغییر به جنسیت، احتمال شرطی  $p(g'|g)$  را محاسبه می کنیم که همسایه تصادفی افراد با جنسیت  $g$  جنسیت  $g'$  داشته باشد که در آن مرد را با  $M$  و زن را با  $F$  نشان می دهیم. برای دوستان مرد کاربران، دریافتیم که  $p(M|M) = 0.5131$  و  $p(F|M) = 0.4869$ . برای دوستان کاربران زن، دریافتیم که  $p(F|F) = 0.5178$  و  $p(M|F) = 0.4822$ . در هر دو مورد، می بینیم که همسایه تصادفی احتمال بیشتری دارد زن باشد برای درک این نتیجه، احتمال پیروی از یک لبه به طور تصادفی انتخاب شده و رسیدن به یک جنسیت خاص را مقایسه می کنیم.

این احتمالات به ترتیب با  $p(F) = 0.5156$  و  $p(M) = 0.4844$  داده می شود. این احتمال برای زن ها بیشتر است، زیرا تعداد انتهای لبه ها که در ادبیات شبکه ها خرد نامیده می شوند، بیشتر از مردان است. در حالی که تقریباً 30 میلیون کاربر زن فعال کمتر در فیس بوک وجود دارد، میانگین مدرک تحصیلی زنان (198) بزرگتر از میانگین مردان است. درجه (172)، که منجر به  $p(F) > p(M)$  می شود. با مقایسه این مقادیر می بینیم که  $p(F|M) < p(F|F)$  و  $p(M|F) < p(M|M)$ . با این حال، بزرگی تفاوت بین این احتمالات بسیار کم است و فقط در مکان هزارم متفاوت است. بنابراین اگر ترجیحی برای دوستی های هم جنسیتی در فیس بوک وجود داشته باشد، تأثیر آن حداقل به نظر می رسد. در نهایت، ما به کشور مبدا می پردازیم، یک متغیر طبقه بندی شده که بر اساس آن به 249 دسته تقسیم می شود.

استاندارد کد کشور ISO 3166-1. این برچسب ها بر اساس آخرین منبع ورود به آدرس IP کاربر و مکاتبات شناخته شده بین آدرس های IP و مکان های جغرافیایی به کاربران نسبت داده می شوند. در حالی که داده های به اصطلاح geo-IP ناقص هستند، عموماً در سطح ملی قابل اعتماد هستند.

به طور شهودی، ما انتظار داریم که دوستان بسیار بیشتری از کشور مبدا خود داشته باشیم تا از خارج از آن کشور، و داده ها نشان می دهد که 84.2 درصد از لبه ها در داخل کشورها هستند. بنابراین شبکه به طور عادلانه تقسیم می شود.

به طور مشخص در امتداد خطوط کشور به خوشه های شبکه یا جوامع. این سازمان در مقیاس مزوسکوپی قابل انتظار است زیرا فیس بوک روابط اجتماعی تقسیم شده بر اساس مرزهای ملی را ثبت می کند. ما می توانیم این تقسیم بندی را با استفاده از مدولاریت  $Q$  [37] که کسری از یال های درون جوامع منهای کسر مورد انتظار یال های درون جوامع در یک نسخه تصادفی شبکه است که درجات را برای هر فرد حفظ می کند، اندازه گیری کنیم [38]، اما در غیر این صورت تصادفی است.

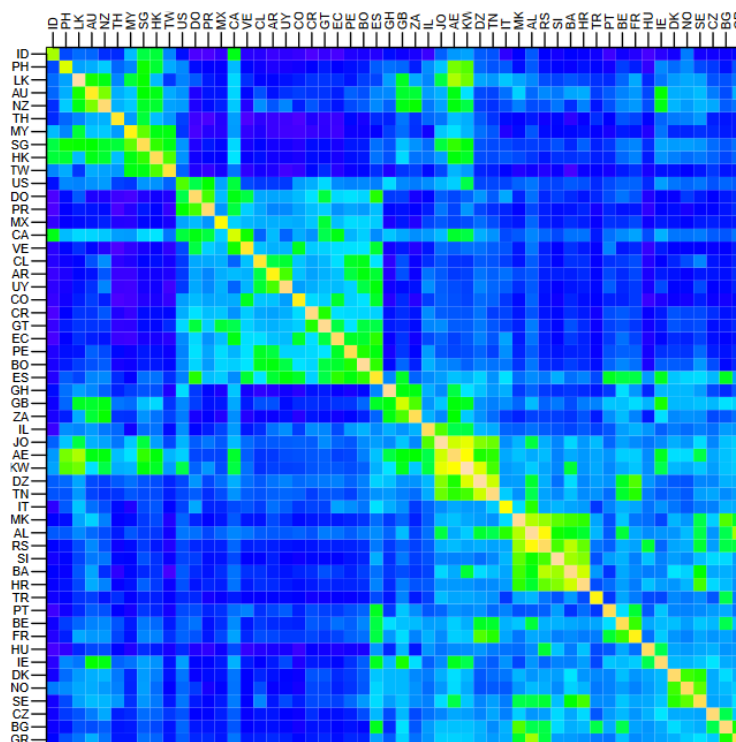


Figure 9. Normalized country adjacency matrix. Matrix of edges between countries with > 1 million users and > 50% Facebook penetration shown on a log scale. To normalize, we divided each element of the adjacency matrix by the product of the row country degree and column country degree.

در این حالت جوامع، کشورها هستند. مقدار محاسبه شده  $Q = 0.7486$  است که بسیار بزرگ است [39] و یک ساختار شبکه به شدت مدولار در مقیاس کشورها را نشان می دهد. به خصوص با توجه به این که برخلاف مطالعات متعددی که از مدولاریت برای شناسایی جوامع استفاده می کنند، ما به هیچ وجه سعی نکردیم مستقیماً آن را به حداکثر برسانیم، و در عوض صرفاً از کشورهای معین به عنوان برچسب های جامعه استفاده کردیم. ما این ساختار بسیار مدولار را در شکل 9 تجسم می کنیم. این شکل یک نقشه حرارتی از تعداد لبه ها بین 54 کشوری را نشان می دهد که جمعیت کاربران فعال فیس بوک از یک میلیون کاربر فراتر می رود و بیش از 50 درصد از جمعیت فعال اینترنت هستند [40]. برای اینکه کاملاً دقیق باشد، ماتریس نشان داده شده حاوی هر یال دو بار، یک بار در هر دو جهت است، و بنابراین دارای دو برابر تعداد یال ها در عناصر مورب است. تعداد یال ها با تقسیم ورودی  $ij$  بر مجموع ردیف و ستون، برابر با حاصل ضرب درجات کشور  $i$  و  $j$  نرمال شد. سپس ترتیب کشورها از طریق خوشه بندی سلسله مراتبی لینک کامل تعیین شد.

در حالی که بیشتر لبه های شکل بصورت مورب قرار دارند، مقیاس ورود به سیستم به وضوح ساختار مدولار بیشتری را در دوستی بین کشورها نشان می دهد. کشورها به گروه هایی تقسیم می شوند، الگوهای کاملاً مربع مانند در ماتریس، با الگوهای دوستی ترجیحی در میان شهروندان کشورهای مختلف. فهرست کامل کشورها به ترتیب ارائه شده در ماتریس در جدول 1 نشان داده شده است. در بسیاری از کشورها گروه بندی بر اساس جغرافیا بصری است.

به عنوان مثال، گروه های واضح مربوط به اقیانوس آرام جنوبی، آمریکای شمالی و مرکزی، آمریکای جنوبی، آفریقای شمالی و خاورمیانه، اروپای شرقی و مدیترانه و به کشورهای شمال اروپا دانمارک، سوئد و نروژ. دیگر گروه بندی های کنجکاوتر، که به وضوح بر اساس جغرافیا نیستند، شامل ترکیبی از بریتانیا، غنا و آفریقای جنوبی، که ممکن است منعکس کننده روابط تاریخی قوی باشد. این شکل به وضوح نشان می دهد که نه تنها دوستی ها عمدتاً بین کاربران در همان کشور است، بلکه دوستی بین کشورها نیز بسیار مرسوم است و ظاهراً تحت تأثیر جغرافیا قرار دارد. تأثیر فاصله جغرافیایی بر دوستی قبلاً مورد بحث قرار گرفته است [41،42]، اما کار قبلی در فیس بوک ساختار دوستی در مقیاس کشور را بررسی نکرده است [43]. حدس و گمان، برخی از گروه بندی های کشورها ممکن است با روابط تاریخی و فرهنگی بهتر از فاصله جغرافیایی ساده توضیح داده شوند.

## بحث

در این مقاله، ساختار نمودار اجتماعی فیس بوک را با استفاده از معیارهای بسیاری مشخص کرده ایم. امیدواریم این اطلاعات هم برای تحقیقات علوم اجتماعی و هم برای طراحی نسل بعدی نمودار الگوریتم ها و تکنیک های تحلیل شبکه های اجتماعی مفید باشد.

مواد و روش ها، محاسبات بر روی یک خوشه Hadoop با 2250 ماشین، با استفاده از چارچوب تجزیه و تحلیل داده Hadoop/Hive در فیس بوک توسعه یافته است [45، 46]. برای تجزیه و تحلیل محله های شبکه، 5000 کاربر به طور تصادفی با استفاده از نمونه گیری مخزن برای هر یک از 100 اندازه محله با فاصله log انتخاب شدند و نمونه ای از 500000 کاربر ایجاد کردند که تجزیه و تحلیل برای آنها انجام شد. صدک های نشان داده شده برای ضریب خوشه بندی و هم تراز برای هر محله است. بنابراین، اندازه صدک های تجربی از این جامعه نمونه 5000 کاربر است. برای تجزیه و تحلیل ساختار اجزای شبکه، از الگوریتم نیومن-زیپف (NZ) استفاده کردیم [47]. الگوریتم NZ، نوعی از الگوریتم Union-Find با فشرده سازی مسیر، ساختار مؤلفه ها را به صورت پویا با اضافه شدن یال ها به شبکه ای که کاملاً خالی از یال ها شروع می شود، ثبت می کند. وقتی همه لبه ها اضافه شده هستند، الگوریتم ساختار اجزای شبکه را محاسبه کرده است. برای اهداف ما، الگوریتم NZ نیازی به حفظ لبه ها در حافظه ندارد. ما آن را در فیس بوک اعمال می کنیم.

شبکه روی یک کامپیوتر با 64 گیگابایت رم با پخش جریانی در لیست لبه ها برای محاسبات طول مسیر، توابع همسایگی بر روی یک ماشین 24 هسته ای با 72 گیگابایت رم با استفاده از الگوریتم HyperANF [48]، به طور میانگین در 10 اجرا محاسبه شدند. برای جزئیات فنی پشت این ابزار محاسبات سنگین، [27] را ببینید. طبق دانش ما، مطالعه ما بزرگترین تحلیل ساختاری یک شبکه اجتماعی است که در تاریخ برای آن انجام شده است. ما با مشخص کردن توزیع درجه شروع کردیم که نشان داده شد با واریانس زیادی در تعداد دوستی ها منحرف است. برخلاف بسیاری از شبکه های دیگر، قانون قدرت خالص برای این شبکه توزیع درجه فیس بوک نامناسب به نظر می رسد، اگرچه هاب ها قطعاً وجود دارند.

سپس اثر دنیای کوچک و شش درجه جدایی در مقیاس واقعاً جهانی تأیید شد. میانگین فاصله بین رئوس مؤلفه غول پیکر 4.7 است و ما این نتیجه را به این صورت تفسیر می کنیم که نشان می دهد افراد در فیس بوک به طور بالقوه



دسترسی فوق العاده ای دارند. محتوای اشتراک گذاری شده فقط باید چند مرحله در شبکه اجتماعی فیسبوک پیش برود تا به بخش قابل توجهی از جمعیت جهان برسد. ما دریافتیم که شبکه اجتماعی فیس بوک تقریباً به طور کامل متصل است، مسیر متوسط طول کوتاهی و خوشه بندی بالا دارد. بسیاری از شبکه های تجربی دیگر، اجتماعی و غیراجتماعی نیز با این ویژگی ها مشاهده شده اند و واتس و استروگاتز شبکه هایی با این ویژگی ها را کوچک می نامند.

شبکه های جهانی [44]. از آنجایی که دوستان ما بیشتر از ما دوستان دارند، افراد در فیس بوک به طرز شگفت انگیزی تعداد زیادی از دوستان دوستان دارند. علاوه بر این، دوستان ما به شدت خوشه بندی شده اند و دوستی های ما دارای هسته های متراکم هستند، پدیده ای که در شبکه های اجتماعی کوچک تر دیده نمی شود. این ساختار همسایگی قابل توجه است. مفاهیم الگوریتمی برای محاسبات پیمایش گراف جستجوی عرض اول تا فاصله دو به طور بالقوه تعداد زیادی از افراد را به دلیل همبستگی درجه مثبت پرس و جو می کند و سپس بسیاری از لبه ها به افرادی که قبلاً به دلیل خوشه بندی پیدا شده اند دنبال می شود.

ما همچنین یک مقایسه اکتشافی از ساختار شبکه با ویژگی های کاربر از جمله رفتار ورود به سیستم، سن، جنسیت و کشور انجام دادیم. ما الگوهای ترکیبی جالبی پیدا کردیم، از جمله اینکه «دوستان شما بیشتر از شما وارد فیس بوک می شوند» و ترجیح قوی برای دوستان هم سن و سال و از یک کشور دارند.

ساختار جامعه به وضوح در شبکه جهانی، در مقیاس دوستی بین کشورها و در درون کشورها، آشکار است. کشورها نیز به نوبه خود یک سازمان مدولار را نشان می دهند که عمده تاً توسط فاصله جغرافیایی مشخص شده است. برخلاف مطالعات قبلی در مورد شبکه ها، این ساختار جامعه بدون تلاش زیاد کشف شد، که نشان دهنده بینش های ساختاری قابل توجهی است که می توان از اطلاعات اولیه جمعیت شناختی به دست آورد. در حالی که محاسبات ما بسیاری از جنبه های ساختار بزرگترین شبکه اجتماعی جهان را روشن کرده است، ما مطمئناً امکانات تحلیل شبکه را تمام نکرده ایم.

تا اینجا ترجمه مقاله را نوشتم و در این قسمت به بررسی تحلیل شبکه اجتماعی Facebook می پردازم. استنباط خودم را از خواندن مقاله بطور خلاصه ذکر می کنم.

در این متن نتایج تحلیل بزرگترین شبکه اجتماعی به نام فیس بوک انجام شده با بیش از 721 میلیون کاربر در سراسر جهان شامل بررسی ویژگیهای مختلف از جمله تعداد کاربران ، دوستی ها، توزیع ها، درجه طول مسی، خوشه بندی و الگوهای ترکیبی می شود. اثبات شده است که شبکه بصورت گراف کاملاً متصل است و 99.91٪ کاربران به Giant component متصل هستند و پدیده شش درجه جدایی در مقیاس جهانی را تایید میکند.

همسایه های نموداری کاربران دارای ساختار متراکم هستند ولی نمودار ساختار فیس بوک بطور کلی پراکنده است. مشاهده شده که دوستان شما دوستان بیشتری نسبت به شما در فیس بوک دارند. تاثیر سن و کشور را بر دوستی ها بررسی کردند. ساختار جامعه جهانی مبتنی بر ملیت را مشاهده کردیم. در حال حاضر نقطه کانونی اینترنت از شبکه ای از اسناد به شبکه ای از افراد تغییر کرده و ساختارهای غول پیکر گرافی در شبکه های اجتماعی کاربران را به یکدیگر متصل کرده است. پارامترهای خوشه بندی ، اثر دنیای کوچک و توزیع ناهمگون دوستان و ساختار جامعه هم در این شبکه مورد مطالعه قرار گرفته است.

این شبکه با بیش از 721 میلیون کاربر دارای 68.7 میلیارد یال دوستی و بطور میانگین هر کاربر 190 دوست فیس بوکی داشته است. یک زیر گراف آن 149 میلیون کاربر آمریکایی دارد. در آمریکا هر کاربر بطور میانگین 214 دوست فیس بوکی دارد که بیشتر از میانگین جهانی است. هاب ، جمعیت کوچکی از کاربران با درجات بالا هستند. تابع توزیع در مقیاس log-log است. 92 درصد از کاربران فیس بوک در پنج درجه جدایی و 99.6 درصد در شش درجه هستند. میانگین ضریب خوشه بندی 14 درصد است که بیانگر این است که یک کاربر متوسط 14 درصد از کل جفتهای دوستشان ، خودشان هم دوست هستند. اثبات شده که اکثر کاربران فیس بوک را با اهداف غیر اجتماعی دنبال می کنند.

با بررسی همبستگی بین درجه تحصیلی یک فرد و لاگین کردن به فیس بوک متوجه میشویم زمانیکه شما در حدود 70 درصد از روزهای یک ماه وارد سیستم شوید دوستان شما بیشتر وارد فیس بوک میشوند. همچنین مشاهده شده 83.6 درصد دوستان کمتری نسبت به تعداد متوسط دوستان خود دارند.

اثبات شده که همبستگی های درجه تعداد دوستی ها در همسایگی شبکه اجتماعی شما به تعداد درجه دوستان و درجه دوستان شما بستگی دارد. میانگین هم ترازوی بعنوان تابعی از درجه کاربر گزارش شده است. در نمودارهای همسایگی کاربران ، میانگین هم ترازوی تابع درجه کاربر در حال افزایش است. مثلاً کاربر با 100 دوست میانگین هم ترازوی 15 و کاربر با 500 دوست میانگین هم ترازوی 53 دارد یعنی حداقل 54 نفر دوستانی که همه 53 نفر دوستان خود را می شناسند. دوستی ها حول هسته های متراکم قابل توجهی متمرکز می شوند به دلیل وجود Giant component که نودهای افراد مشهور در آن قرار گرفته اند .

بطور کلی افراد جوان بیشتر دوستان با محدوده سنی کوچک دارند در حالی که افراد مسن تر دارای طیف بسیار گسترده تری هستند. با توجه به ویژگیهای سنی کاربران گراف فیس بوک یک همسایه تصادفی به احتمال زیاد هم

سن شماس. علاوه بر آن , احتمال دوستی با افراد مسن تر به سرعت و تقریبا بطور مستقل از حالت خارج می شوند که تقریبا مستقل از سن کاربر است.

در مورد ویژگی جغرافیایی , داده ها نشان میدهد که 84.2 درصد از لبه ها در داخل کشورها هستند. داده ها به وضوح نشان میدهند که نه تنها دوستی ها عمدتا بین کاربران در همان کشور است بلکه دوستی بین کشورها هم بسیار مرسوم است و تحت تاثیر جغرافیا قرار دارد. همین طور , میانگین فاصله بین رئوس مولفه گول پیکر گراف 4.7 است. نتیجه می گیریم که افراد در فیس بوک بطور بالقوه دسترسی فوق العاده دارند.