[آسان داک](http://www.asandoc.com/) (www.Asandoc.com)

**ارزیابی ریسک محرمانگی (پنهانی) در شبکه های اجتماعی آنلاین**

**چکیده**

سنجش ریسک محرمانگی در شبکه های اجتماعی آنلاین، یک کار چالش بر انگیز است. یکی از مشکلات اساسی، سنجش مقدار اطلاعات افشا شده به صورت ناخواسته و غیر عمد است. ما PrivAware را معرفی می کنیم که ابزاری برای تشخیص و گزارش از دست رفتن اطلاعات ناخواسته در شبکه های اجتماعی آنلاین است. هدف ما، فراهم کردن یک چارچوب اصولی برای شناسایی ریسک محرمانگی و فراهم کردن راه حل هایی برای کاهش از دست رفتن اطلاعات است. اولین گام نرم افزار بر از دست رفتن اطلاعات مرتبط با چرخه های اجتماعی تمرکز می کند. در نوشته های بعدی ما می خواهیم قابلیت های اضافی برای به دست آوردنمدل های فرعی تهدید و خطر را یکپارچه کنیم. با توجه به نتایج اولیه، ما ریسک محرمانگی مرتبط با روابط دوستی در فیسبوک را مورد سنجش قرار می دهیم. نشان می دهیم که برای هر کاربر در مطالعه ی ما، اکثریت خصوصیات شخصیشان می تواند از روابط اجتماعی ناشی شود. علاوه بر این، ما نتایجی را ارائه می کنیم که حاکی از آن است که تعداد دوستان به یک خصوصیت به وجود آمدهمرتبط هستند. ما همچنین نتایج مشابهی را برای جمعیت شناسی های مختلف کاربران ارائه می کنیم. هدف PrivAware، نه تنها گزارش اطلاعات از دست رفته است بلکه اقداماتی را برای کاهش ریسک محرمانگی به کاربران پیشنهاد می دهد. اقدامات، برای کاربران، مراحل ضروری را برای بهبود کل سنجش محرمانگی خود، ارائه می دهند. همانطور که آشکار است اما ایده آل نیست، راه حل، حذف دوستان مخاطره آمیز است. روش دیگر، گروه بندی دوستان مخاطره آمیز و به کارگیری کنترل های دستیابی به گروه برای محدود کردن قابل رویت است. به طور خلاصه، هدف ما، فراهم کردن یک ابزار منحصر به فرد برای سنجش از دست رفتن اطلاعات و فراهم کردن ویژگی هایی برای کاهش ریسک محرمانگی است.

**1.مقدمه**

شبکه های اجتماعی آنلاین، در چند سال اخیر خیلی پر طرفدار شده اند. از این نوشته، فیسسوک بیش از 200 میلیون کاربر فعال دارد و بیش از 100 میلیون کاربر حداقل یک بار در روز در فیسبوک وارد حساب کاربری خود می شوند (1). شبکه های اجتماعی آنلاین، بسترهایی را برای کاربران خود فراهم می کنند تا اطلاعات شخصی خود را همگانی کنند.

یک اصل اساسی در امنیت کامپیوتر، ممانعت اطلاعات از فرار از مرزهای محرمانگی در نظر گرفته شده است. به اطلاعاتی که فراتر از قسمت های تعریف شده توسعه می یابند به طور معمول نشت اطلاعات نامیده می شود. در بیشتر محیط ها، سنجش مقدار اطلاعات از دست رفته، یک کاردشوار است. علاوه بر این، یکپارچه سازی اطلاعات از دست رفته با خطر خاص حتی می تواند چالش بر انگیز تر باشد. ما قصد داریم اطلاعات نشان داده شده به طور ناخواسته را در شبکه های اجتماعی آنلاین بسنجیم و راه حل هایی برای کاهش ریسک محرمانگی ارائه کنیم.

اهمیت سنجش محرمانگی در شبکه های اجتماعی آنلاین؛ با توجه به مقیاس شبکه ها، حتی حیاتی تر است. حفاظت از مقدار عظیمی از اطلاعات شخصی متناظر، یک کار مهم و حیاتی است. مثال های اخیر (2)، (3)، مکانیسم های کنونی را مطرح می کنند که سطوح نامناسبی از حفاظت را ارائه می کنند. علاوه بر این، مقاله ی اخیر (4)، مطرح می کند که کاربران مشتاقریسک از دست دادن کنترل اطلاعات شخصیشان هستند. به منظور ارزیابی ریسک های محرمانگی مرتبط با شبکه های اجتماعی، ما اول خطرات و تهدیدات مختلف را شناسایی کرده و می سنجیم.

تحقیق ما بر سنجش ریسک های محرمانگی در شبکه های اجتماعی آنلاین و فراهم کردن راه حل هایی برای کاهش آن ریسک ها تمرکز می کنیم. برای کمک به سنجش تهدیدات محرمانگی، PrivAware ، ابزاری برای سنجش ریسک محرمانگی در فیسبوک را معرفی می کنیم. PrivAwareبرای اجرا در پروفایل یک کاربر برای فراهم کردن گزارش دهی و مجموعه ای از اقدامات پیشنهادی برای کاهش تهدیدات محرمانگی طراحی شده است. این نوشته ی کنونی، ریسک محرمانگی مرتبط با روابط دوستانه در فیسبوک را می سنجد. به علاوه، نوشته، راه حل های ساده ای را برای ریسک محرمانگی مرتبط با این تهدید فراهم می کند.

93 شرکت کننده برای نصب و اجرای PrivAwareانتخاب شدند. به طور میانگین، توانستیم 59.5% از خصوصیات شخصی مرتبط با شرکت کنندگانمان را ریشه یابی و استنتاج کنیم. برای تمام جمعیت شناختی ها، زن، مرد، متاهل، و غیر متاهل، توانستیم خصوصیات شخصی مرتبط با کاربران را بیش از 50% اوقاتاستنتاج کنیم. علاوه بر سنجش ها همچنین اقدامات کاربران را برای شرکت کنندگانمان برای کمک به کاهش ریسکشان فراهم کردیم. یک اقدام، حذف روابط دوستیپر مخاطره است که منجر به از دست رفتن اطلاعات می شود. دیگری، یک رویکرد موشکافانه تر، جداسازی دوستان متخطی به گروه ها و به کارگیری مکانیسم های کنترل دستیابی برای هر گروه است. برای مثال در فیسبوک، یک کاربر می تواند کنترل های دسترسی را برای محدود کردن کارکرد مرتبط با یک گروه خاص از کاربران به کار ببرد. PrivAwareگروه بندی کاربران پر مخاطره را به جای حذف آنها پیشنهاد می دهد. برای فراهم کردن لیست دوستان مخاطره آمیز، ما مجموعه هایی از روش های آروینی برای محدود کردن کل ریسک محرمانگی در کنار حداکثر سازی تعداد روابط دوستان را به کار می گیریم. روش های آروینی می توانستند بهبود های چشمگیری بر یک رویکرد خط مبنافراهم کنند که به طور راندوم کاربران را حذف می کند تا جایی که سطح مورد نظر محرمانگی ایجاد گردد. به طور میانگین، تعداد دوستان مورد نیاز برای حذف یا گروه بندی، با استفاده از روش آروینی متداول دوستان، 19، کمتر از خط مبنا بود.

ما می خواهیم قابلیت های کنونی PrivAware را برای در بر گرفتن مکانیسم های بعدی برای سنجش و گزارش مدل های تهدید اضافی توسعه دهیم. هدف از این نتایج اولیه، جلب توجه به نیاز به سنجش ریسک محرمانگی در شبکه های اجتماعی و تشویق تحقیق بعدی است.

باقیمانده ی این مقاله به صورت زیر سازماندهی شده است. بخش 2، بخش کارهای مرتبط است. بخش 3، جزییات طراحی را برای PrivAware ارائه می کند. بخش 4، تجربه را توصیف می کند. بخش 5 و 6، به ترتیب، کار آتی و نتیجه گیری می باشد.

**2.کارهای مرتبط**

مقیاس شبکه های اجتماعی آنلاین در کنار افزایش مراقبت، بر عاملان شبکه (فیسبوک، MySpace و غیره) فشار آورده تا سطوح افزایش یافته ای از محرمانگی کاربر را فراهم سازند. به طور خاص، مطالعات اخیر از برخی نگرانی های محرمانگی چالش بر انگیز پرده برداری کرده اند (5)، (6)، (7). اگرچه، تلاش های کمی، بازخورد قابل ملاحظه ای را برای کاربر نهایی فراهم کرده است. تحقیق ما مربوط به شناسایی و سنجش تهدیدات محرمانگی برای فراهم کردن گزارش دهی می باشد بنابراین کاربران نهایی می توانند بسته به سطح مورد نظر خود از ریسک محرمانگی، تصمیم گیری کنند.

در حال حاضر، PrivAware محدود به سنجش ریسک محرمانگی مرتبط با روابط اجتماعی مستقیم می شود. تحقیق ما، در کشف و بررسی این پدیده، اولین نیست. تجزیه و تحلیل روابط اجتماعی برای نشان دادن خصوصیات کاربر، در تحقیق پیشین تست شده است (8). در این کار، نویسندگان، تکنیک های یادگیری آماری مختلف را برای اشاره ی صحیح به ارزش برای یک نوع خصوصیت خاص به کار برده اند. برای مثال، در مطالعه ی آنها، نویسندگان می توانستند به طور صحیح به مقادیر خصوصیت را از دید سیاسی و جنسیت در فیسبوک اشاره کنند. به طور میانگین، بیشتر از 50 % اوقات. ما این نتایج را برای کارمان، تمجید آمیز می بینیم و داده های اضافی را برای تقویت موقعیتمان ارائه می کنیم.

کار مشابهی همچنین در زمینه کشفاستنباط و نتیجه گیری انجام شده است. در این نمونه، محققان یک ابزار نیمه اتوماتیک برای ایجاد اطلاعات خصوصی از منابع داده ی تدوین شده، توسعه دادند. تشخیص استنباط، با استفاده از تجزیه و تحلیل تکرر عبارت در کنار نتایج جستجوی آنلاین انجام می شود. با یک ابزار، نویسندگان می توانستند نسبت بزرگی از اطلاعات خصوصی از مستندات تدوین شده را پوشش دهند. نویسندگان همچنین صورت گرایی جدید را برای سنجش اطلاعاتی که به طور ناخواسته نشان داده شده اند را ارائه کنند. این کار برای تصمیم ما برای دنبال کردن سنجش ریسک مخاطره آمیز در شبکه های اجتماعی آنلاین اثر می گذارد. در (9) محققان مشکلات در اعتباربخشی نتایجشان را تجربه کردند. مجموعه داده های کمی موجود هستند که شامل هم مستندات تدوین شده و هم اصلی می شوند. در دامنه ی شبکه های اجتماعی، ما محدود به چنین چیزی نیستیم. می توانیم نتایج استنباط خود را به طور مستقیم با استفاده از گره های هدف اطلاعات اصلی پروفایل اثبات کنیم.

تحقیق ما، از این کارها و کارهای زیاد دیگر، بهره می برد. اگرچه، ما عقیده داریم که کارمان، کمک های جدیدی به این زمینه می کند. اول، ما خصوصیات زیادی را در نظر می گیریم. تحقیقمان را محدود به یک مجموعه ی داده ی ساختار یافته نمی کنیم. برای مثال، ما استنباط ها را بر اساس مقادیر نیمه ساختار یافته مثل کارکنان و ساختار سازمانی ایجاد می کنیم. در نتیجه، ما مجبور شدیم به مشکلات چالش بر انگیز مثل ابهام زدایی (10) و تشخیص ماهیت (11) اشاره کنیم. دوم، راه حل هایی را برای کمک به کاهش نتایج از یافته هایمان ارائه کردیم. ما الگوریتم های مختلف را برای کاهش از دست رفتن اطلاعات در نظر گرفتیم، پیچیدگی های زمان اجرایمتناظر را بررسی کردیم و اقداماتی را برای کاربران برای کاهش ریسک محرمانگی شان مطرح کردیم. در پایان، و از همه مهم تر، ما چارچوبی را مطرح می کنیم که به تحقیق بعدی برای توسعه کارمان جهت در برگرفتن کارهای اضافی برای مدل های تهدید مختلف توسعه دهد.

**3. طراحی**

برای شناسایی و کاهش ریسک های محرمانگی مرتبط با دوستان در شبکه های اجتماعی آنلاین، ما PrivAware را پیشنهاد می کنیم. PrivAware، دو کار را فراهم می کند: اول، خصوصیات یک کاربر را بر اساس خصوصیات دوستانش ایجاد می کند. دوم، پیشنهاد می دهد که چگونه برای کاهش تعداد خصوصیات به وجود آمده به یک سطح قابل قبول، اعضای دوستان کاربر را تغییر داد.

**3.1. تشخیص نتیجه گیری**

ما به طور معمول، مشکل نتیجه گیری را به صورت زیر تعریف می کنیم. کاربر t، هدف نتیجه گیری می باشد. F هم مجموعه ای از دوستان مستقیم t می باشد. مشکل نتیجه گیری این است: با توجه به تمام خصوصیات کاربران در F، خصوصیات tبه وجود می آید.

**3.1.1.محاسبات استنباط.**این مشکل، چندین مشکل عینی را در شبکه های اجتماعی نشان می دهد. برای مثال، یک کاربر آگاه از محرمانگی t، چندین خصوصیت خصوصی را تنها برای گروه قابل دسترسی می گذارد. یک عضو خارج از گروه ممکن است هنوز بتواند مقادیر خصوصیات قابل دسترسی گروه را بر اساس مقادیر دوستان tایجاد کند. در مثال دیگری، یک کاربر که محرمانگی ندارد، یک برنامه ی شبکه اجتماعی نصب کرده است. از آنجایی که آن برنامه می تواند به تمام خصوصیاتش دسترسی پیدا کند، انها به طور هدفمند خصوصیات مشخصی را خالی می گذارد. اگرچه، از آنجایی که برنامه ممکن است به خصوصیات بدون محرمانگی دوستان کاربر دسترسی یابد، ممکن است بتواند خصوصیات از قلم افتاده ی کاربر را ایجاد کند.

بر اساس مشاهدات کهچرخه های اجتماعی گرایش به افشای اطلاعات دارند، ما خصوصیات یک کاربر را بر اساس خصوصیات دوستانشان ایجاد می کنیم. برخی ممکن است ضرب المثل، کبوتر با کبوتر باز با باز را به یادآورند. ما از یک الگوریتم ساده و شهودی برای نشان دادن قدرت این تکنیک برای به خدمت گرفتن خط مبنا برای مقایسه استفاده می کنیم. برای هر ویژگی و خصوصیت، الگوریتم، رایج ترین مقدار این خصوصیت را در میان دوستان کاربر انتخاب می کند. اگر تعداد دوستانی که این ارزش را به اشتراک گذاشته اند فراتر از آستانه باشد، الگوریتم این ارزش را به عنوان ویژگی ایجاد شده ی کاربر تعیین می کند. در حال حاضر، PrivAware، برداشت هایی را برای خصوصیت زیر ریشه یابی کرده است: سن، کشور، ایالت، کد پستی، نام دبیرستان، سال فارغ التحصیلی از دبیرستان، دانشگاه، مدرک، کارفرما، نسبت، وضعیت رابطه، و دیدگاه سیاسی.

**3.1.2. ابهام زدایی.**الگوریتم ساده که در بالا توضیح داده شده با یک چالش روبرو می شود: اگر دو خصوصیت مساوی باشند، وقتی ارزش مفهومی مشابهی به طور متفاوت نشان داده شوند، چگونه تشخیص داده شوند؟ برای مثال UC Berkely، Berkeley و Cal، متغیرهای دانشگاه کالیفرنیا، Berkeley هستند. این به طور متداول به یک مشکل ابهام زدایی داده مربوط می شود که معمولا در رابطه با متن نیمه ساخت یافته و غیر ساخت یافته است. برای کاهش این مشکل، ما تعدادی رویکرد به کار می بریم. اول، یک فرهنگ از متغیرهای متداول برای دانشگاه، طرفین سیاسی، مدارک، و کارفرمایان توسعه دادیم. PrivAware، از فرهنگ ها به عنوان یک جدول جستجو برای انتقال مقادیر به اشکال متعارف استفاده می کند. برای مثال، PrivAwareارزش خصوصیت دکترای فلسفه را به Ph.D و Cal را به UC Berkeleyمنتقل می کند. دوم، هر مقدار را با استفاده از الگوریتم لونشتین(Levenshtein) (12) پردازش می کنیم. این رویکرد، عبارات متشکل از اشتباه املایی و تفاوت های نقطه گذاری را یکسان می کند.

**3.1.3. تایید.** برای ارزیابی قدرت الگوریتم ساده ی استنباط، از معیارهای زیر استفاده می کنیم:

خصوصیات به وجود آمده: خصوصیاتی که PrivAware می تواند ایجاد کند. وقتی تعداد دوستانی که یک ارزش خصوصیت مشابه را کمتر از یک آستانه به اشتراک می گذارند،PrivAwareنمی توان یک خصوصیت را ایجاد کند.

استنباط های قابل اثبات: خصوصیاتی که PrivAware می تواند ایجاد کند و همچنین در پروفایل کاربر هدف نشان داده می شوند. در این روش، ما صحت خصوصیات ایجاد شده را اثبات می کنیم.

استنباط های صحیح: خصوصیاتی که PrivAware می تواند ایجاد کند که با ارزش در پروفایل کاربر هدف جور است.

**3.2. کاهش استنباط**

وقتی PrivAware نشان می دهد که خصوصیات خصوصی مشخص یا خصوصیات غایب یک کاربر ممکن است ایجاد شوند، کاربر ممکن است بخواهد مجموعه ای از دوستانش را برای جلوگیری از ایجاد تغییر دهد. برای مغلوب شدن الگوریتم ایجاد PrivAware، کاربر نمی تواند دو استراتژی اتخاذ کند. اول، کاربر می تواند دوستان را حذف کند. وقتی تعداد دوستانی که خصوصیات مشابهی را به اشتراک می گذارند که به زیر آستانه برسد، الگوریتم ایجاد PrivAwareشکست می خورد. در حقیقت، کاربر دو انتخاب دارد. او می تواند به سادگی دوستان پر مخاطره را همانطور که PrivAware شناسایی کرده حذف کند (بعدا مورد بحث قرار خواهد گرفت). یا، اگر بستر شبکه ی اجتماعی او از کنترل دستیابی در گروه ها پشتیبانی می کند (مثل فیسبوک)، او می تواند دوستانش را به گروه های ایمن و غیر ایمن تقسیم کند، گروه های نا امن را غیر قاب لرویت کند و تمام دوستان پر مخاطره ای که PrivAware شناسایی کرده را به گروه نا امن ببرد (مثلا این دوستان را از دید عمومی مخفی کند). روش دیگر، اغلب مطلوب تر است چون کاربران بعید است که بخواهند دوستان را حذف کنند به ویژه آنهایی که ارزش های خصوصیتی مشابه دارند. به جای حذف یا مخفی کردن دوستان، کاربر می تواند شبکه اش را از دوستان با اضافه کردن دوستان تقلبی اشغال و شلوغ کند. برای نمونه، اگر کاربر به اندازه ی کافی دوستان تقلبی اضافه کرده باشد که متداول ترین ارزش یک ویژگی در میان دوستان تقلبی باشد، الگوریتم استنباط ما این ویژگی را خروجی می گیرد که با خصوصیت درست کاربر همخوانی ندارد. اگرچه این روش، عیب است. اول، کاربر نباید دوستان تقلبی اضافه کند چون این کار دوستان واقعی اش را گیج می کند و شبکه های اجتماعی اش را غیر عادی جلوه می دهد. دوم، دوستان تقلبی نباید درخواست های اضافه کردن را بپذیرند که باید مانع از کاربران برای اضافه کردن این دوست های تقلبی در بسترهای شبکه اجتماعی خاص شوند. ما فقط اولین روش را بررسی خواهیم کرد به عبارت دیگر شناسایی دوستان پرمخاطره برای حذف یا مخفی کردن.

**3.2.1. تعریف مشکل.**ما مشکلمان را به صورت زیر تعریف می کنیم. ما دوستان مرتبط با کاربر t را به عنوان مجموعه ای چند تایینشان می دهیم . عبارت ، مقدار تخصیص یافته به دوست را توسط کاربر t نشان می دهد. این عبارت مفید است چون به ما در بهینه سازی نتایج در توافق با دوستانی که ارزش اجتماعی بالاتر دارند، اجازه می دهد. برای مثال، یک کاربر باید یک ارزش اجتماعی بالاتر به دوستانی اختصاص دهد که اعضای خانواده هستند نسبت به دوستانی که همکلاسی هستند. ارزش دیگر در مجموعه چند تایی، مجموعه ی متناظری از خصوصیات مرتبط با دوست است. ما هر ویژگی در مجموعه را سه تایی نشان می دهیم. کلمه ی نوع، نوع خصوصیت مرتبط با ارزش است مثل دانشگاه، سن، کد پستی. به طور مشابه، کلمه ی ارزش متناظر با ارزش واقعی برای یک خصوصیت خاص است مثل UC Berkeley، 25، 95812. کلمه ی نهایی، وزن، ارزشی در دامنه ی صفر تا یک است که بر اساس اعتماد فرآیند ابهام زدایی تخصیص می یابد. برای مثال، خدمات ابهام زدایی باید یک امتیاز 0.8 برای مقدار خصوصیت Cal و یک امتیاز 1 برای UC Berkeley اختصاص دهد. الگوریتم استنباط ما از این وزن برای محاسبه تکرر مقادیر خصوصیت استفاده می کند. برای مثال، اگر دانشگاه، برای یک دوست Cal و برای دوست دیگر، UC Berkeley باشد، آنها 0.8+1.0=1.8 برای مقدار تکرر مقدار خصوصیت متعارف UC Berkeley می شود. جدول 1، خصوصیات یک دوست را نشان می دهد.



جدول 1. خصوصیات یک دوست

به توجه به توصیف بالا از دوستان، هدف ما، کاهش تعداد خصوصیات به وجود آمده توسط حذف یا مخفی کردن دوستان است. ظاهرا، هرچه دوستان بیشتری را حذف کنیم، تعداد کمتری از خصوصیات را ایجاد می کنیم. بر اما از سوی دیگر، کاربر می خواهد تا جای ممکن دوستان زیادی را تحت الزامات محرمانگی حفظ کند. بنابراین، ما مشکل کاهش استنباط را به صورت زیر تعریف می کنیم:

تحت الزامات محرمانگی، به عنوان ماکزیمم تعداد اجازه داده شده خصوصیات قابل استنباط، و مجموعه ای از دوستان، زیر مجموعه ای از دوستان را بیابید که مقادیر کل دوستان را به حداکثر می رساند و اینکه تحت الزامات محرمانگی است.

**3.2.2. راه حل های آروینی (کاوشی).**ما روش های مبتنی بر راه های آروینی را برای کاهش ریسک محرمانگی ادامه می دهیم. همچنین یک نمونه ی خاص تر از مشکل را در نظر می گیریم. ما ارزش تمام دوستان را برابر می کنیم. همچنین ابهام زدایی کاملی را فرض می کنیم، بنابراین وزن برای هر خصوصیت، ثابت است. با انجام این دو تغییر، می توانیم راه حل های مهار پذیری را برای حداقل سازی ریسک محرمانگی فراهم کنیم.

حذف تصادفی دوستان. این الگوریتم، راست و مستقیم است و به عنوان خط مبنا برای مقایسه به کار می رود. آن حذف تصادفی دوستان را ادامه می دهد تا زمانی که به الزام محرمانگی برسد (تعداد مجاز ماکزیمم خصوصیات به وجود آمده).

حذف دوستان با بیشترین خصوصیات. حذف تصادفی دوستان شکست می خورد تا تفاوت بین دوستان در اثرهایشان براستنباط در نظر گرفته شود. به طور شهودی، ما می خواهیم دوستانی را حذف کنیم که به استنباط بیشتر از سایرین اثر می گذارند. اولین روش آروینی، در طول هر تکرار است، که دوستی را با بیشترین تعداد خصوصیات قابل رویت حذف می کند. این، بر اساس شهود است که دوستان با تعداد زیاد تری از خصوصیات با مقدار بزرگ تری برای تشخیص استنباط تاثیر می گذارد.

حذف دوستان با مشترک ترین دوستان. الگوریتم دوم آروینی ما، تعداد دوستان مشترک را بین کاربر هدف و هر یک از دوستانش در نظر می گیرد. در طول هر تکرار، الگوریتم، دوستی که مشترک ترین دوستان را با کاربر هدف به اشتراک می گذارد، حذف می کند. شهود این است که افرادی که دوستان بیشتری را به اشتراک می گذارند، همچنین خصوصیات مشترک تری را به اشتراک می گذارند.

**3.2.3. ملاحظات طراحی.**PrivAware، برای اجرای هم در شبکه های OpenSocialو هم در فیسبوک طراحی شده بود. اگرچه ما فیسبوک را برای گردآوری نتایج مقدماتیمان به دو دلیل اصلی انتخاب کردیم. دلیل اول و مشهود تر، مقدار داده های موجود است. 200 میلیون کاربر هم اکنون در فیسبوک ثبت نام کرده اند. این، فرصت را برای PrivAwareفراهم می کند تا به طور ویروسی از طریق شبکه گسترش یابد (نه به یک روش بدجنسی) و به ما فرصتی می دهد تا به مقدار عظیمی از محتوای ایجاد شده ی کاربران نفوذ کنیم. مزیت عمده ی دیگر، یک تفاوت ریز در محرمانگی بین فیسبوک و OpenSocialوجود دارد. در OpenSocial، یک برنامه ی ثالث می تواند تنها اگر هر دو طرفین (کاربر و دوست) راضی باشند و برنامه را نصب کرده باشند، داده ی دوست کاربر را جستجو کند (13). به عبارت دیگر، اگر کاربر الف PrivAware را در شبکه OpenSocial، نصب و اجرا کند، برنامه تنها اطلاعات دوست را برای دوستان کاربر الف که PrivAware را هم نصب کرده اند؛ جستجو می کند. در مقابل، فیسبوک این محدودیت را ندارد. اگر کاربر الف، PrivAware را در فیسبوک اجرا کند، داده های مرتبط با دوستان کاربر الف را جستجو می کند. این تفاوت به ما اجازه می دهد تا اطلاعات دوست را برای یک کاربر که PrivAware را اجرا می کند، گردآوری و بررسی کنیم.

**4.تجارب**

**4.1. گردآوری داده**

ما PrivAware را برای سنجش ریسک محرمانگی برای یک تهدید محرمانگی معروف در مورد برنامه های ثالث توسعه دادیم (14). هدف PrivAware، تعیین این بود که اگر سنجش محرمانگی نامساعدی وجود داشته باشد آیا کاربران اقدام خواهند کرد یا نه. برای گمراه کردن شرکت کنندگان، ما یا تبلیغ در فیسبوک با متن زیر قرار دادیم.

درباره محرمانگی؟ PrivAware، ابزاری برای سنجش قابلیت دستیابی اطلاعات پروفایل برای برنامه های کاربردی است. تعیین کنید که چه مقدار از اطلاعاتتان را به نمایش می گذارید.

اگر یک کاربر بر تبلیغ کلیک کند، آنها به صفحه ی خانگی PrivAware منتقل می شدند که شامل توضیح زیر بود.

PrivAware یک برنامه ی کاربردی ساده ی فیسبوک است که برای امتیاز دادن به تنظیمات محرمانگی در مورد برنامه های کاربردی ثالث طراحی شده است. آن پروفایل شما را بررسی خواهد کرد تا تعیین کند کدام اطلاعات برای برنامه های کاربردی فیسبوک موجود هستند. بر اساس مقدار اطلاعات موجود، PrivAwareدرجه ی متناظر را تخصیص می دهد. علاوه بر امتیاز، به کاربران مجموعه ای از سوالات در مورد محرمانگی یادآوری می شود. بعد از کامل کردن پرسشنامه، کاربران تشویق می شوند تا تنظیمات محرمانگی خود را دوباره پیکربندی کنند و امتیاز محرمانگی خود را دوباره محاسبه کنند. سوالات در بازدیدهای بعدی برای کاربر یادآوری نخواهد شد.

ما 105 نفر را دریافتیم که مشتاق شرکت در مطالعه هستند. از هر شرکت کننده در مطالعه یمان خواستیم تا به مجموعه ای از سوالات قبل و بعد از نشان دادن امتیاز محرمانگی، پاسخ دهند. هدف، جلب تمایل شرکت کنندگان وقتی که در معرض ریسک محرمانگی خود قرار می گرفتند، بود. نتایج پرسشنامه در جدول 2 لیست شده اند. نتایج به درصد به استثنای امتیازهای محرمانگی کاربر داده شده است، آنها به عنوان میانگین ارائه شده اند.



جدول 2. واکنش ها به تهدید محرمانگی

برای محاسبه ریسک محرمانگی، ما به سادگی تعداد کل خصوصیات قابل رویت را به برنامه های کاربردی ثالث بر تعداد کل خصوصیات به ازای هر شرکت کننده تقسیم کردیم. نتیجه ی محاسبه شده، درصد خصوصیات نشان داده شده به برنامه های کاربردی ثالث است. برای سادگی وقتی که به کاربر نشان داده می شود، ما درصدها را به یک درجه حروف از A تا Fتغییر می دادیم. امتیاز F متناظر با تعداد بزرگی از خصوصیات نشان داده شده است و امتیاز A، تعداد کمی از خصوصیات نشان داده شده را نشان می دهد. توزیع امتیازها برای 105 شرکت کننده به صورت زیر است: 64% برای F، 36% برای D، و هیچ شرکت کننده ای به A، B و Cامتیاز ندادند. این نتایج مطرح می کنند که ریسک محرمانگی مربوط به تهدید برنامه کاربردی ثالث بدجنس، بالا است.

برای پاسخ به سوالات ابتدایی ما، آیا کاربران برای کاهش سطح بالای ریسک محرمانگی اقدام می کنند، ما نتایج پرسشنامه را در زمینه ی امتیازهای محرمانگی بررسی کردیم. قبل از اینکه درجه به شرکت کننده نشان داده شود از آنها خواسته می شود تا به سه سوال پاسخ دهند. سوال اول تعین آشنایی آنها با سیاست محرمانگی فیسبوک در مورد برنامه های کاربردی ثالث بود. با توجه به نتایج، 51.4% احساس می کردند که کمی با این سیاست آشنا هستند. این مطرح می کند که شرکت کنندگان ما، در مورد ریسک محرمانگی در نظر دارند. بعدی، از شرکت کنندگان خواسته شد که چگونه به تنظیمات محرمانگی خود از 1 تا 5 امتیاز می دهند. 1 بیانگر هیچ محرمانگی است و 5 نشان دهنده ی محرمانگی کامل است. امتیاز میانگین، 3.057 بود. این حاکی از آن است که کاربران گمان می کنند که بر اساس تنظیمات محرمانگی کنونی خود، نیمی از اطلاعاتشان برای برنامه های کاربردی ثالث در دسترس باشد. در پایان، از شرکت کنندگان سوال شد که آیااز هر گونه مکانیسم محرمانگی مهیا شده توسط فیسبوک استفاده می کنند. به طور شگفت انگیز، 100% از هیچ مکانیسمی استفاده نمی کردند. بر اساس این نتایج، ما می توانیم شرکت کنندگان خود را از مدافعان محرمانگی میانه تا ضعیف دسته بندی کنیم. وقتی شرکت کننده به مجموعه ی اول سوالات پاسخ داد، امتیاز ریسک محرمانگی شان تعیین شد. برای تکرار، 64% F را امتیاز دادند. در نتیجه، شرکت کنندگان با دو سوال پایانی روبرو شدند. سوال اول دوباره درباره امتیاز محرمانگی پرسید، این زمان تنها کاربر یک درجه مشخص کرد. همانطور که انتظار می رفت، نتایج خیلی پایین تر بودند، امتیاز، 1.047 بود. سوال دوم پرسیده بود که آیا شرکت کنندگان برای تغییر تنظیمات محرمانگی شان اقدام خواهند کرد یا نه. از این نتایج، 64.7% پاسخ دادند که تنظیماتشان را تغییر خواهند داد.

در میان 105 نفر که پرسشنامه ما را تکمیل کردند، 93 نفر برای شرکت در تحقیق تشخیص و کاهش استنباط موافق بودند. جمعیت شناختی شرکت کنندگان حالا به این صورت است. میانگین سن، 23.89 با انحراف استاندارد 6.1 و دامنه ی 14-44 بود. شامل 47 مرد و 24 زن بود. گروه شامل 12 کشور زادگاه مختلف بود: کانادا، چین، اکوادور، مصر، ایران، مالزی، نیوزلند، پاکستان، سنگاپور، آفریقای جنوبی، انگلستان، و ایالات متحده. برخی از شرکت کنندگان، سن، جنسیت، یا تبار خود را بیان نکردند. برای استنتاج امتیازهای محرمانگی مان، تمام 12523 رابطه ی دوستی مستقیم را بررسی کردیم. جدول 3، تعداد کل و میانگین دوستان را نشان می دهد.



جدول 3. تعداد کل و میانگین دوستان

**4.2. تشخیص استنباط**

برای ارزیابی اثربخشی الگوریتم استنباطمان، از معیارهای تعریف شده در بخش 3.1.3 استفاده می کنیم: ویژگی های به وجود آمده، استنباط های قابل اثبات، و استنباط های صحیح. جدول 5، درصد استنباط های صحیح را (در تمام استنباط های قابل اثبات) با هر ویژگی نشان می دهد. نشان می دهد که ویژگی های ساخت یافته (مثل، سن، کشور، ایالت، سال فارغ التحصیلی از دبیرستان) به طور صحیح به درصد بالاتری از زمان گرایش دارند. یک استثنا، کد پستی است. برعکس، ویژگی های نیمه ساخت یافته و غیر ساخت یافته، سخت تر گرایش به به وجود آمدن دارند. با ابهام زدایی داده های بهبود یافته، ما حدس می زنیم که درصد استنباط های صحیح افزایش یابند. انتظار ما، بر اساس نمونه گیری داده ها از نتایجمان است. با تجزیه و تحلیل دستی داده ها، دریافتیم که بسیاری از نمونه ها در حقیقت برابر بودند اما به تنهایی توسط تکنیک های ابهام زدایی داده هایمان، شناسایی نشدند.

جدول 4، تعداد ویژگی های به وجود آمده، استنباط های قابل اثبات، و استنباط های صحیح را یک به یک ذکر کرده است. به علاوه، ما نتایج را برای جمعیت شناختی های مختلف برای شناسایی گرایشات در داده ها، به حساب آوردیم.



جدول 4. کل ویژگی های ایجاد شده



جدول 5. استنباط های صحیح توسط ویژگی

جدول 6، تعداد شرکت کنندگان را برای استنباط های ایجاد شده نشان می دهد. ما یک شرکت کننده را به عنوان دوستی تعریف می کنیم که حداقل در گردآوری ارزش های ایجاد شده، یک کمک فراهم کند. برای مثال، اگر الگوریتم استنباط ما، مقدار دانشگاه استنفورد را برای دانشگاه به وجود آورد، تمام دوستان با دانشگاه استنفورد در پروفایلشان به آن استنباط اثر می گذارند. جدول 7، تعداد میانگین شرکت کنندگان را به ازای هر استنباط، استنباط های قابل اثبات و استنباط های صحیح ارائه می کند.

نتایج بالا برای شرکت کنندگان مطرح می کند که فیسبوک وقتی با تهدید شمارش تکرار ویژگی روبرو می شود، در بهترین حالت، محرمانگی را کمتر از 50 درصد اوقاتارائه می کند.

**4.3.کاهش استنباط**

ما سه الگوریتم را برای حذف دوستان در بخش 3.2.2 به کار بردیم. اول، الگوریتمی را اجرا کردیم که دوستان را به طور تصادفی از تمام تنظیمات شرکت کنندگان حذف کردیم تا جایی که سطح مطلوب محرمانگی به صفر برسد، که محرمانگی کامل را نشان می دهد. تعداد میانگین دوستان باید حذف شوند تا به این سطح برسند، 145 بود. بر اساس اولین ملاحظات، این نتیجه به نظر می رسد که بعید است تعداد میانگین دوستان، 134 باشد. اگرچه، تجزیه و تحلیل داده ها، گرایشی را نشان می دهد که توسط شرکت کنندگان با تعداد زیادی از دوستان نشان داده شده است. برای مثال، شرکت کنندگان با تعدادی از دوستانی بیشتر از پانصد تا، مستلزم تعداد زیاد تری از دوستان هستند تا نسبت به کسانی که دوستان کمتری دارند، حذف کنند. برای کاهش این انحراف، مجموعه ای از داده ها را به گروه هایی از شرکت کنندگان با تعداد افزایش یافته ی دوستان تقسیم بندی کردیم. مجموعه های در نظر گرفته شده در نتایج، به گروه های 50، 100، 200 و 500 دوستان تقسیم شدند.



جدول 6. کل شرکت کنندگان برای استنباط



جدول 7. میانگین شرکت کنندگان برای استنباط



جدول 8. دوستانی که باید حذف شوند

ما همچنین نتیجه ی تمام شرکت کنندگان را برای تکمیل در نظر گرفتیم. جدول 8، نتایج را یک به یک ذکر می کند.

جدول 8 همچنین نتایج را از دو الگوریتم دیگر نشان می دهد یعنی حذف دوستان با بیشترین ویژگی ها و حذف دوستان با بیشترین دوستان مشترک. نتایج نشان می دهند که آخرین الگوریتم (حذف دوستان با بیشترین دوستان مشترک) بهتر از بقیه عمل می کند (منجر به حذف کمترین تعداد دوستان می شود)، در حالی که اولین الگوریتم (حذف تصادفی دوستان) بدترین است. در حقیقت، به طور میانگین، با حذف دوستان مشترک، تفاوت در تعداد دوستان باید حذف شوند یا گروه بندی شوند، در مقابل با رویکرد تصادفی، 19 در 5 گروه کلی بود.

هر دوی این الگوریتم ها، پیشرفت هایی را در حذف دوستان به طور تصادفی برای محدود کردن از دست رفتن محرمانگی فراهم می کنند. اگرچه، هیچ یک از روش ها، ترتیب را به حساب نیاورده اند. همانطور که قبلا اشاره شد، راه حل بهینه، راه حلی است که هر تغییر اساسی را در نظر می گیرد و به حساب می آورد که چگونه ترتیب بر امتیاز محرمانگی اثر می گذارد. برای مثال، حذف زودتر یک دوست در فرآیند باید منجر به سطوح بالاتری از محرمانگی نسبت به حذف در مراحل بعدی در فرآیند شود. علاوه بر این، هیچ یک از این رویکرد ها سعی نمی کنند مشکل کلی تری را حل کنند که نتایج ارزش اجتماعی و ابهام زدایی ناقصرا در نظر می گیرد.

**5.کار آتی**

در حال حاضر، PrivAware، ریسک محرمانگی را برای یک مدل تهدید مجزا در فیسبوک می سنجد. در مطالب آینده، هدف ما، یکپارچه سازی مدل های تهدید مختلف بسیار برای انجام یک ارزیابی کامل تر از ریسک محرمانگی در شبکه های اجتماعی آنلاین می باشد. ما همچنین می خواهیم یک نسخه از PrivAware را اجرا کنیم که در شبکه های OpenSocial اجرا می شود. با چندین متغیر در شبکه های مختلف، ما قابلیت مقایسه و همسنجی ریسک محرمانگی در رابطه با هر شبکه را خواهیم داشت. از این رو، کاربران اطلاعاتی درباره سنجش خواهند داشت که شبکه ی اجتماعی، سطح مطلوبی از محرمانگی را برای مقاومت ریسکشان ارائه می کند. به علاوه، ما تمایل به سنجش ریسک مرتبط با اقدامات متداول کاربر در شبکه های اجتماعی آنلاین داریم. برای مثال، سنجش ریسک در رابطه با تگ کردن آنلاین عکس دوستان یا کاربران در کامنت ها.

بهبود الگوریتم های استفاده شده برای سنجش ریسک محرمانگی، همچنین حوزه ی عمده ی تحقیق است. رویکردها و روش های ارائه شده در اینجا تنها به عنوان یک خط مبنابرای نشان دادند ریسک در مورد روابط دوستی در فیسبوک به کار می روند. هدف ما، فراهم کردن الگوریتم های پیچیده تر برای سنجش بهتر سطح محرمانگی برای چندین مدل تهدید است. ما توسط تحقیق در پردازش زبان طبیعی و داده کاوری تشویق شده ایم تا توسعه ی PrivAware را گسترش دهیم. به ویژه، تکنیک های خوشه بندی داده و الگوریتم های تشخیص ماهیت پیشرفته تر، ممکن است سنجش های محرمانگی بهبود یافته ای را فراهم کند.

جنبه ی مهم دیگر تحقیق ما، ترویج یک چارچوب متداول برای سنجش ریسک محرمانگی در برنامه های کاربردی آنلاین است. هدف غیر صریح و محض تحقیق ما، فراهم کردن مثالی از ایجاد محرمانگی در شبکه های اجتماعی آنلاین است. اگرچه، این کار محدود به شبکه های اجتماعی نمی شود اما در حوزه های دیگر، کاربرد دارد. برای مثال، سنجش ریسک محرمانگی مرتبط با ایمیل آنلاین یا برنامه های کاربردی میکرو بلاگینگ. در آینده می خواهیم تحقیقمان را به این حوزه ها توسعه دهیم و یک سنجش متداول از محرمانگی در این حوزه های مختلف ارائه کنیم.

**6. نتیجه گیری**

سنجش ریسک محرمانگی در شبکه های اجتماعی آنلاین، یک کار مهم است. میلیون ها کاربر با حجم زیادی از اطلاعات برنمودار اجتماعیشان اثر می گذارند. اطلاعاتی که به طور ناخواسته در معرض نمایش قرار داده شده اند، می توانند پیامدهای جدی و خطیری داشته باشند. برای مسائل پیچیده، بسیاری از کاربران با ریسک های اساسی محرمانگی در رابطه با شبکه های اجتماعی آشنا نیستند. اقدامات متداول کاربر مثل اضافه کردن یک دوست می تواند سطح نمایش ناخواسته ی اطلاعات را افزایش دهد.

PrivAware قصد دارد مقدار اطلاعات نمایش داده شده در شبکه های اجتماعی آنلاین را ارزیابی کند و روش هایی را برای کاهش آن ریسک ها ارائه کند. در این متن کنونی، ما از دست دادن اطلاعات در رابطه با رابطه ی دوستی در فیسبوک را بررسی کردیم. با توجه به نتایج ما، 59.5% اوقات توانستیم ویژگی های یک کابر را بر اساس ارتباطات اجتماعیشان به دست آوریم. همچنین نتایجی را برای جمعیت شناختی های مختلف کاربران فراهم کردیم که ویژگی هایی را مطرح می کنند که می توان با احتمال بزرگ تر از 50% اوقات به وجود آورد و استنباط کرد. علاوه بر گزارش دهی ریسک محرمانگی، توانستیم اقدامات کاربر را برای شرکت کننده جهت کمک به کاهش ریسک محرمانگیشان تامین کنیم. به طور میانگین، تعداد دوستانی که باید برای محرمانگی کامل حذف شوند یا گروه بندی شوند، با استفاده از روش آروینی دوستان مشترک، 19 تا کمتر از خط مبنا بود. این نتایج امیدوار کننده هستند و کاتالیزوری (حلالی) برای تحقیق آتی فراهم می کند. خدف بلند مدت ما، ارائه ی سیستمی است که چندین مدل تهدید را می سنجد و برای کاربران، جهت کاهش ریسک های محرمانگی شان، راهنمایی ارائه می کند.

**منابع**