# استنتاج اعتماد در شبکه‌های اجتماعی ناهمگون

**مسعود آخوندی[1]، جعفر حبیبی[2]، شهروز معاون[3]**

[1]دانشجوی کارشناسی ارشد، فناوری اطلاعات، دانشگاه صنعتی شریف

akhoondi@ce.sharif.edu

[2]دانشیار، گروه نرم‌افزار و فناوری اطلاعات، دانشگاه صنعتی شریف

jhabibi@sharif.edu

[3] دانشجوی کارشناسی ارشد، نرم‌افزار، دانشگاه صنعتی شریف

moaven@ce.sharif.edu

## ۱- مقدمه

شبکه‌های اجتماعی از تاریخچه‌ای طولانی در شاخه‌های مختلف علوم برخوردار می‌باشند. در سال ۱۹۶۰ میلادی، Milgram فرضیه‌ای به نام «*جهان خرد*» را ارائه کرد، در این فرضیه افراد مختلف در جهان با تعداد کمی از آشنایانشان به هم وصل می‌شوند که این تعداد در تئوری شش نفر بود [10]. در [11] نشان داده شده است که شبکه‌های اجتماعی بسیاری وجود دارند که ویژگی‌های مشترک جهان خرد را به اشتراک می‌گذارند و در این شبکه‌ها میانگین فاصله بین گره‌ها کم بوده و از ضریب ارتباطی خوشه‌بندی بالایی برخوردار می‌باشند. در علوم کامپیوتر شبکه‌های اجتماعی برای نشان‌دادن و مدل‌کردن «*رابطه*» بین اشیا مختلف استفاده شده است و هر نوع سیستم پیچیده‌ای مانند اینترنت را می‌توان با این شبکه‌ها نمایش داد. در [12] نشان داده شده است که وب نیز به عنوان یک شبکه اجتماعی می‌تواند به گونه‌ای مناسب خوشه-بندی شود و طول مسیر بین گره‌های آن (وب سایت‌ها) کم می‌باشد.

رشد شبکه‌های اجتماعی وبی و ویژگی‌های این شبکه‌ها در مدل‌کردن مسائل مختلف، پتانسیل فراوانی را برای تولید نرم‌افزارهای هوشمند فراهم کرده است. یکی از کاربردهای این شبکه‌ها برای مدل‌کردن و استنتاج اعتماد i می‌باشد. استفاده از اعتماد و شهرت i i در زندگی روزمره امری مهم و مرسوم است. به همین علت اعتماد در شاخه‌های مختلف علوم نظیر جامعه شناسی [1]، روانشناسی [2] و اقتصاد [3] از جنبه‌های گوناگون بررسی شده است. در زمینه علوم کامپیوتر و فناوری اطلاعات و ارتباطات، مطالعه اعتماد و شهرت کاربردهای فراوانی داشته است؛ به خصوص درحوزه تجارت الکترونیک از سیستم‌های اعتماد و شهرت به طور گسترده استفاده می‌شود. از همین رو از اعتماد در زمینه‌های مختلف علوم با توجه به زمینه کاربرد، تعاریف گوناگونی وجود دارد. Gambetta، اعتماد را به عنوان قابلیت اطمینان یک فرد یا یک چیز، به صورت زیر تعریف می‌کند [4]: «اعتماد احتمالی برای عنصر A است که عنصر A از یک عنصر دیگر به نام B انتظار دارد یک عمل معین را که منفعت وی را دربر دارد انجام دهد». Mui تعریف دیگری از اعتماد دارد که بر مبنای شهرت بنا شده است [5]: «اعتماد، انتظاری فردی یک عامل در مورد رفتار آینده یک عامل دیگر بر اساس تاریخچه رویارویی با آن عامل تعریف می‌شود».

بحث دیگری که امروزه در علوم کامپیوتر مطرح است، استخراج رابطه از روی رابطه‌های موجود می‌باشد. فرآیند ترکیب‌کردن چندین رابطه برای شکل دادن یک رابطه جدید بر مبنای دانش قبلی در مورد رابطه هدف، « استخراج رابطه i i i» نامیده می‌شود. این دانش قبلی می‌تواند به شکل یک عبارت پرس‌وجوی ساده توسط کاربر فراهم شود. استخراج رابطه به عنوان مثال می‌تواند به طور گسترده در وب‌سایت‌هایی با انجمن-های مجازی برای پیشنهاد دوست، بازاریابی و پیش‌بینی نمودن استفاده شود[13]. در [13] برای استخراج رابطه از روی رابطه‌ای خطی بر روابط پایه تکیه می‌کند اما ما در این مقاله راه‌حلی برای استفاده از عملگرهای فازی گسترش می‌دهیم و نشان می‌دهیم؛ استفاده از این عملگرها می-تواند نتیجه بهتری را به همراه داشته باشد.

روشهای موجود برای استنتاج اعتماد با استفاده از شبکه‌های اجتماعی، فرض می‌کنند که این شبکه‌ها همگن بوده و تمامی لینک‌ها و نیز گره‌های شبکه از یک نوع می‌باشند؛ به عبارت دیگر تمایزی میان آنها وجود ندارد [6، 7، 8، 9]. موارد فراوانی وجود دارد که این فرض در مورد آنها درست نیست. در روش پیشنهادی ما با در نظر گرفتن نیاز کاربران (با استفاده از استخراج رابطه مناسب) خروجی متناسب با این نیازها را تولید می‌نماییم. به عنوان مثال یک فروشگاه الکترونیکی را در نظر بگیرید که افراد در آن برای کالاهای مختلف میزانی از اعتماد را تعیین می-

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.
www.SID.ir

Archive of SID

کنند. در این حالت هر کالا و نظرات افراد در مورد آن، بیانگر یک شبکه اجتماعی (یک رابطه) است. در روش پیشنهادی ما در هنگام انجام یک معامله جدید که در آن تعدادی از افراد مشارکت دارند، ابتدا به استخراج روابطی که همه یا برخی از این افراد در آنها با هم تعامل داشته، پرداخته و سپس با تشکیل یک رابطه هدف i با استفاده از ترکیبی از روابط استخراج شده، به استنتاج اعتماد می‌پردازیم.

ساختار این مقاله چنین است که ابتدا در قسمت ۲، به مرور کارهای مرتبط در دو حوزه اعتماد و استخراج رابطه پرداخته می‌شود. در قسمت ۳ روش‌های پیشنهادی ما برای استنتاج اعتماد آورده شده است. قسمت ۴ دربرگیرنده تحلیل روش‌های پیشنهاد شده است. نتایج تجربی حاصل از شبیه‌سازی نیز در قسمت ۵ آورده شده و در نهایت نیز در قسمت ۶ به جمع‌بندی و نتیجه‌گیری مقاله پرداخته می‌شود.

## ۲- کارهای مرتبط

در این قسمت ابتدا به مرور کارهای مرتبط در حوزه اعتماد پرداخته می‌شود و برخی از مهمترین روشهای مورد استفاده در زمینه استنتاج اعتماد معرفی شده است. در بخش بعدی، مروری اجمالی به کارهای انجام‌شده در زمینه استخراج رابطه آورده شده است.

### ۲-۱- اعتماد

Marsh در [21] یک روش بنیادی برای اعتماد ارائه کرده است که مبنای بسیاری از روش‌های اعتماد دیگر می‌باشد. این روش پیشنهادی تنها اعتماد میان دو گره را مدل می‌کند و به علت پیچیدگی بیش از اندازه، بندرت در عمل استفاده می‌شود.

در مدیریت اعتمادی که الگوریتم Marsh معرفی می‌کند، به توصیه‌های سایر گره‌ها پرداخته نمی‌شود و تنها اعتماد بین دو گره را بررسی می‌کند.

یکی دیگر از روش‌های مطرح برای استنتاج اعتماد، روش Tidal است که در [26] آمده است. این روش به طور کامل در قسمت ۳-۴ معرفی شده است.

روش ارائه‌شده در [22] جهت استفاده در انجمن‌های مجازی برای استفاده در تجارت الکترونیکی و عامل[۷]های خودمختار مصنوعی بوجود آمد و با اعتماد، همانند تصور انسان به همان‌شکلی که در جوامع واقعی است، برخورد می‌کند. این روش از دو نوع اعتماد، «اعتماد مستقیم» و «اعتمادِ پیشنهاددهنده» استفاده می‌کند. اعتماد مستقیم به اعتماد یک عامل به دیگری بر اساس تجربه مستقیم اطلاق می‌شود؛ اعتمادِ پیشنهاددهنده نیز به اعتماد یک عامل به توانایی دیگری در فراهم آوردن پیشنهادات خوب، گفته می‌شود. مقادیری که برای اعتماد توسط گره‌ها داده می‌شود، گسسته بوده و برای اعتماد مستقیم عبارتند از «خیلی قابل اعتماد، قابل اعتماد و غیر قابل اعتماد» و برای اعتماد پیشنهاد دهنده «خیلی خوب، خوب؛ بد و خیلی بد» می‌باشند. مزیت اصلی در این الگوریتم، توانایی غلبه بر این موضوع است که عامل‌های مختلف ممکن است از سطوح مشابهی از اعتماد استفاده کنند ولی هر کدام برداشت متفاوتی از این سطوح داشته باشند. این تفاوت را می‌توان به عنوان «فاصله معنایی» محاسبه نمود.

این روش همچنین توانایی کار با عدم قطعیت را دارد؛ بدین معنا که در صورتی که یک عامل نتواند مقدار اعتمادی را برای یک عامل دیگر مشخص نماید (مثلا به همان اندازه‌ای تجربه خوب داشته باشد که تجربه خیلی خوب دارد)، این الگوریتم از شیوه‌های خاصی برای حل موضوع بهره می‌برد. با این وجود چگونگی سود جستن از عدم قطعیت در فرآیند محاسبه اعتماد، روشن و صریح نیست.

در این روش برای محاسبه مقدار اعتماد، پیشنهادهای عامل‌های ناشناس را درنظر نمی‌گیرد، اما این عامل‌ها به عنوان فراهم آورندگان پیشنهاد شناخته شده و پیشنهادهای بعدی آنها به عنوان بخشی از محاسبات بعدی مورد استفاده قرار می‌گیرد. جمع آوری داده‌ها، فقط از عامل‌هایی که با عاملِ هدف تجربه مستقیم دارند، انجام می‌گیرد.

روش بعدی روش Regret است. این روش با اعتماد در محیط‌های کوچک و متوسط در حوزه تجارت الکترونیک سروکار دارد. اطلاعات مورد نیاز از روابط میان بخش‌های تعاملی و محیط بدست می‌آید. که این روابط می‌توانند رقابتی، مشارکتی یا تبادلی باشند[23,24,25] .

این روش سه بعد برای اعتماد در نظر می‌گیرد. بعد فردی که بر اساس تجربیات شخصی یک فرد است که "اعتماد مستقیم" نامیده می‌شود. بعد اجتماعی بر اساس اطلاعات شخص سوم، ارتباطات میان افراد و نقش اجتماعی افراد است. بعد سوم، بعد آنتولوژیکی است که به انتقال اطلاعات مربوط به اعتماد در بین زمینه‌های مرتبط کمک می‌کند.

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.
www.SID.ir

## ۲-۲- استخراج روابط

داده کاوی انجمن [vi] از شبکه‌های اجتماعی، در واقع تقسیم گره‌های شبکه به دسته‌هایی با ارتباطات درون انجمنی متراکم و بین انجمنی کم، می‌باشد. شناسایی این ساختارها به فهم و مصورسازی ساختار شبکه‌های اجتماعی کمک می‌کند [14، 15، 16]. مطالعه این ساختارها با پارتیشن‌کردن گراف در تئوری گراف و خوشه‌بندی سلسله مراتبی در جامعه‌شناسی در ارتباط است[14، 17، 18].

الگوریتم‌های مرسوم برای داده‌کاوی انجمن فرض می‌کنند که تنها یک نوع ارتباط بین گره‌ها وجود دارد و تنها تعداد کمی از روشهای پیشنهاد شده برای استخراج رابطه و داده‌کاوی انجمن‌ها ارتباط‌های گوناگونی بین اشیا در نظر می‌گیرند که به شبکه‌هایی با چنین ویژگی، شبکه‌های اجتماعی ناهمگن یا چند رابطه‌ای گفته می‌شود [13].

یک راه‌حل برای استخراج رابطه در چنین شبکه‌هایی، استفاده از ترکیبات رابطه‌های مختلف با توجه به تمایلات کاربر، به منظور شکل‌دهی یک رابطه (استخراج رابطه) و سپس استفاده از یک الگوریتم مرسوم برای استخراج رابطه می‌باشد. برای مثال یک جامعه انسانی را در نظر بگیرید که روابط زیادی بین افراد آن برقرار است، این روابط می‌توانند رفتن به یک دانشگاه، گوش کردن به یک نوع موسیقی، داشتن جنسیت یکسان و داشتن بیمه از یک کمپانی باشند. حال فرض کنید که یک وب‌سایت می‌خواهد یک سرویس پیشنهاد دوست به مشتریانش ارائه کند؛ کاربر چند نفر از افرادی را که می‌داند با یکدیگر دوست هستند به سیستم معرفی می‌کند. الگوریتم تلاش می‌کند تا تاثیر هر یک از روابط پایه را بر روی رابطه خواسته شده بیان کند. مثلا یک خروجی این الگوریتم می‌تواند این باشد که رفتن به یک دانشگاه و گوش‌دادن به یک نوع موسیقی تاثیر زیادی در رابطه خواسته شده دارد، جنسیت تاثیر کمی دارد و داشتن بیمه از یک کمپانی اصلا تاثیری ندارد. بر مبنای این دانش و روابط پایه، الگوریتم می‌تواند رابطه‌ای را بسازد که به عنوان یک رابطه دوستی در میان افراد پیش‌بینی می‌کند.

## ۳- استنتاج اعتماد

در این قسمت، ابتدا مساله استخراج رابطه را به صورت فرمال مطابق [13] تعریف می‌کنیم؛ آن را گسترش داده و در نهایت، دو الگوریتم ژنتیک برای تخمین زدن این مساله ارائه می‌دهیم.

### ۳-۱- تعریف مساله

مساله استخراج رابطه در [13] بدین شکل آورده شده است:

«یک مجموعه داده شده از اشیا و یک مجموعه از روابط می‌توانند با یک مجموعه از گرافها به شکل $G_i(V, E_i)$، $i=1,\ldots,n$ نمایش داده شوند که n تعداد روابط، V مجموعه گره‌ها (اشیا) و $E_i$ مجموعه یال‌های مربوط به iامین رابطه است. وزن‌های روی یال‌ها نشان‌دهنده اهمیت ارتباط بین دو شی است. ما از نماد $M_i$ برای مشخص کردن ماتریس وزن‌دار مرتبط با $G_i$، $i= 1,\ldots,n$ استفاده می‌کنیم. فرض کنید یک رابطه مخفی گره‌ها وجود دارد که با $G'(V,E')$ نمایش داده می‌شود و $M'$ ماتریس وزن‌دار متناسب با رابطه $G'$ را مشخص می‌کند. یک مجموعه از اشیا نامدار $X=[x_1,x_2,\ldots,x_m]$ و $Y=[y_1,y_2,\ldots,y_m]$ داده شده است به گونه‌ای که $y_i$ برچسب $x_i$ است (چنین اشیا برچسب‌داری مشخص کننده اطلاعات جزئی از رابطه مخفی است). هدف پیدا کردن یک ترکیب خطی از ماتریسهای وزن‌دار است که بتواند بهترین تخمین از ماتریس مخفی $M'$ را بدست آورد».

یک عبارت پرس‌وجو یا در شکل بردارهای X و Y یا به شکل ماتریس $m\times m$ به نام Q که این ماتریس از X و Y چنین تعریف می‌شود: $Q_{ij}=1$ اگر $x_i$ و $x_j$ دارای برچسب مشابه ($y_i=y_j$) باشند و در غیر اینصورت صفر است. ما از عبارت «*ماتریس هدف*»، برای مشخص کردن Q استفاده می‌کنیم.

در برخی مسائل ممکن است تعداد روابط پایه بسیار باشد یا اینکه هنگام استخراج رابطه، مقدار ضرایب برخی از این رابطه‌ها ناچیز و کم اهمیت باشد که در نتیجه کاربر تنها خواهان بدست آوردن تعداد مشخصی از مهمترین و تاثیرگذارترین روابط در رابطه هدف باشد. در قسمت ۴ در این مورد بیشتر توضیح داده شده است. با توجه به این فرض، مساله استخراج را گسترش داده و به صورت زیر تعریف می‌کنیم:

$M_i$، $G_i$، $G'$ و $M'$ مانند قبل تعریف می‌شوند. اکنون مساله، پیداکردن K تا رابطه‌ای است که ترکیب آنها، بهترین تخمین از ماتریس مخفی $M'$ را بدست می‌دهد. هرمعیاری از شباهت می‌تواند در این مسائل استفاده شود که ما بازهم از همان Frobenius norm اختلاف دو ماتریس استفاده می‌کنیم. بنابراین ما می‌خواهیم دنباله‌ای از ماتریس‌های $I=(i_1,i_2,\ldots i_k)$ وضرایب $A=(a_1,a_2,\ldots a_k)$ که Frobenius norm بین اختلاف M و Q را کمینه می‌کند، پیدا کنیم به طوریکه: $M=\sum_{j=1}^{k} a_j . M_{i_j}$ .

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.
www.SID.ir

Frobenius norm یک ماتریس mxn، B به صورت زیر تعریف می‌شود:

$$||B||_f = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|b_{ij}|^2} \quad (1)$$

در استخراج رابطه، از ماتریس‌های پایه، تنها اطلاعات مورد علاقه ما، زیر قسمتی از ماتریس‌هاست که با انتخاب اشیا در X مشخص می‌شود. بنابراین ما سایر اشیا را در نظر نمی‌گیریم، زیرا آنها در عبارت پرس‌وجوی کاربر وجود ندارند. در نتیجه در این قسمت ما، $M_i$ را به منظور مشخص کردن ماتریس m×m مرتبط با $G_i'$ ($G_i$ زیرگرافی از X) به جای ماتریس n×n مرتبط با $G_i$ در نظر می‌گیریم. به طور مشابه به جای کارکردن با M' می‌توانیم فقط قسمت شناخته شده Q را در نظر بگیریم.

## ۳-۲- الگوریتم ۱

الگوریتم‌های ژنتیک توسط سیستم‌های بیولوژیکی معرفی شدند و سعی دارند fitness راه‌حل‌ها را از طریق سیر تکاملی بهبود ببخشند [19]. یک راه‌حل برای یک مساله به صورت یک رشته به نام کروموزم کد می‌شود که شامل یک مجموعه از عناصر به نام ژن است که خود این ژن‌ها، یک مجموعه مقادیر برای متغییرهایی که قرار است بهینه شوند، نگهداری می‌کنند [20].

از تصمیمات مهمی که در ارتباط با الگوریتم ژنتیک باید گرفته‌شود، نحوه کدکردن مساله می‌باشد. در هر دو الگورتیم، هر راه‌حل، دنباله‌ای nتایی، $S=(s_1,s_2,\ldots,s_n)$ است که در آن iامین عنصر، ضریب ماتریس $M_i$ را مشخص می‌کند. هر $s_i$ می‌تواند مقداری بین صفر تا یک داشته باشد. بنابراین هر فرد S در الگوریتم یک راه‌حل به شکل $\sum_{i=1}^{n} s_i.M_i$ تولید می‌کند. ما با انتخاب یک آستانه روی تعداد ضرایب غیر صفر در هر راه‌حل، فضای جستجو را محدود می‌کنیم. هر راه‌حل اجازه دارد که حداکثر K تا ضریب غیرصفر داشته باشد، به طوریکه K یک مقدار ثابت از قبل تعریف شده است. حال عملگرهای mutation و crossover را به گونه‌ای که با این محدودیت مساله سازگار باشد، تعریف می‌کنیم.

دو راه‌حل $S_1=(s_{11},s_{21},\ldots,s_{n1})$ و $S_2=(s_{12},s_{22},\ldots,s_{n2})$ و راه‌حل $S$ که خروجی عملگر crossover بر روی $S_1$ و $S_2$ می‌باشد را در نظر بگیرید. هر یک از این راه‌حلها، اجازه دارد که حداکثر Kتا ضریب غیرصفر داشته باشد. ما نیاز داریم که ضرایب غیرصفر S را از $S_1$ و $S_2$ انتخاب کنیم. بدون تغییر در مساله فرض می‌کنیم تعداد ضرایب غیرصفر $s_1$ و $s_2$ را به ترتیب با $k_1$ و $k_2$ نشان دهیم و $NZ1=\{s_{i_1 1}, s_{i_2 1}, \ldots, s_{i_{k_1} 1}\}$ و $NZ2=\{s_{j_1 2}, s_{j_2 2}, \ldots, s_{j_{k_1} 2}\}$ نیز به ترتیب مجموعه ضرایب‌های غیرصفر $s_1$ و $s_2$ باشند. یک عدد رندوم $\alpha$ بین صفر تا یک را انتخاب کرده و $S=S_1 \oplus_\alpha S_2$ را چنین تعریف می‌کنیم: $\alpha k_1$ ضریب را از $NZ_1$-$NZ_2$ انتخاب کرده و مقدار آنها را برابر $s_i=s_{i1}$ قرار می‌دهیم. سپس $(1-\alpha)k_2$ ضریب را از $NZ_2$-$NZ_1$ انتخاب کرده و $s_i=s_{i2}$ قرار می‌دهیم. برای همه $i \in NZ_1 \cap NZ_2$ نیز مقدار $s_i$ را برابر $s_i=\alpha.s_{i1}+(1-\alpha)s_{i2}$ قرار می‌دهیم. سایر ضرایب صفر می‌باشند. به عبارت دیگر برای اینکه S، حداکثر K تا ضریب غیر صفر داشته باشد ما $\alpha$*۱۰۰ درصد از ضرایب را در مکانهایی که $s_1$ غیرصفر و $s_2$ صفر است، از $s_1$ انتخاب می‌کنیم، سپس $100(1-\alpha)$ درصد از ضرایب را در مکانی که $s_2$ غیرصفر و $s_1$ صفر است، انتخاب می‌کنیم و نهایتاً همه مکانهای i که هر دوی $s_{i1}$ و $s_{i2}$ غیرصفر هستند، انتخاب می‌شود. واضح است که اگر $k_1$ و $k_2$ کمتر از K باشند، تعداد ضرایب غیرصفر در S نیز کمتر از K خواهد بود.

برای عملگر mutation نیز باید در هر راه‌حل تعداد ضرایب غیرصفر را کمتر از K نگه دارد. اگر تعداد ضرایب غیر صفر کمتر از Kتا باشد که یک ضریب را به طور رندوم انتخاب کرده و مقدار آن را به صورت رندوم تغییر می‌دهیم. در غیر اینصورت یک مکان با ضریب غیرصفر انتخاب شده و مقدار آن تغییر می‌کند (ممکن است تبدیل به صفر شود).

اگر Ms ماتریس مرتبط با راه‌حل S، $M_S = \sum_{i=1}^{n} s_i.M_i$ باشد. fitness هر راه‌حل، مجددا از طریق Frobenius norm محاسبه می‌شود که برای دو ماتریس Ms و Q برابر است با: $\sum_{i=1}^{n}\sum_{j=1}^{n}(m_{s_{ij}} - q_{ij})^2$.

## ۳-۳- الگوریتم ۲

در این الگوریتم نیز کدگذاری، crossover و mutation مشابه الگوریتم ۱ است. تنها تفاوت در تابع fitness می‌باشد. در اینجا ما از عملگر فازی XOR هم استفاده می‌کنیم و fitness هر راه‌حل را بدین صورت محاسبه می‌کنیم: برای هر راه‌حل $S=(s_1,s_2,\ldots,s_n)$ و ماتریس مربوط به آن، $M_s$، داریم $M_S = XOR_{i=1}^{n} a_i.M_i$. با استفاده از این تعریف، fitness یک راه‌حل چنین تعریف می‌شود: $fitness(S)=\sum_{i=1}^{n}\sum_{j=1}^{n}(m_{s_{ij}} - q_{ij})^2$. اکنون ما تنها نیاز داریم عملیات فازی XOR بین دو ماتریس را تعریف کنیم، که این عملیات را به صورت XOR بیت به بیت درایه‌های دو ماتریس تعریف می‌کنیم.

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.
www.SID.ir

### ۳-۴- الگوریتم Tidal

در [26] الگوریتمی برای استنباط اعتماد با استفاده از شبکه‌های اجتماعی مطرح شده است. هنگامی که دو فرد در یک شبکه به طور مستقیم به یکدیگر متصل هستند، مقدار اعتماد به طور صریح وجود دارد. اما در صورتی که بین افراد مسیر مستقیمی وجود نداشته باشد، در صورتی که مسیری از طریق سایر افراد، وجود داشته باشد، می‌توان مقدار اعتماد را با تقریب خوبی بدست آورد. Golbeck نشان می‌دهد که در صورت انتخاب مسیر کوتاهتر بین دو فرد، جواب دقیق‌تری بدست می‌آید. همچنین نشان می‌دهد که در صورت انتخاب مسیرهای با مقدار اعتماد بالاتر، حاصل دقیق‌تر خواهد بود. به عبارت دیگر در صورت وجود چند مسیر با طول یکسان، مسیری که دارای مقدار اعتماد بیشتری است، نتیجه بهتری را به همراه خواهد داشت. الگوریتم پیشنهادی [26] به صورت جستجوی سطح اول، شبکه را پیمایش می‌کند. مقدار اعتماد در این الگوریتم از فرمول زیر بدست می‌آید:

$$t_{is}=\frac{\sum_{j\in adj(j)|t_{ij}\geq max} t_{ij}t_{js}}{\sum_{j\in adj(j)|t_{ij}\geq max} t_{ij}} \qquad (۲)$$

که در آن $t_{ij}$ مقدار اعتماد فرد i به فرد j است و adj(i)، همسایگان i را مشخص می‌کند.

### ۳-۵- استنتاج

هنگامی که کاربر درخواست محاسبه اعتماد میان برخی از گره‌ها را به الگوریتم می‌دهد؛ الگوریتم ابتدا به استخراج رابطه با توجه به پرس‌وجوی کاربر و تعداد رابطه‌های دلخواه وی پرداخته و نتیجه اولیه را که همان رابطه هدف است، برمی‌گرداند. در صورتی که در رابطه هدف افراد شرکت کننده در پرس‌وجو به طور مستقیم به یکدیگر متصل باشند، همان رابطه به عنوان خروجی نهایی، ارائه خواهد شد، در غیر اینصورت با استفاده از یک الگوریتم استنتاج اعتماد که در این مقاله الگوریتم Tidal پیشنهاد شده است، به استنتاج اعتماد بین تمامی افراد پرس‌وجو می‌پردازیم.

## ۴- تحلیل و مقایسه

در این قسمت به تحلیل و مقایسه روش پیشنهادی استخراج رابطه با شیوه پیشنهادشده در [13] می‌پردازیم.

برای استخراج رابطه [13] یک الگوریتم با استفاده از رگرسیون و یک الگوریتم با استفاده از min-cut پیشنهاد می‌دهد. این الگوریتم‌ها برای پیدا کردن راه‌حل دقیق مساله به کار می‌روند (البته در حالتی که مساله، راه‌حل دقیق داشته باشد). با این وجود هر یک از این دو الگوریتم محدودیتهایی دارند که برای برخی از سناریوها قابل اعمال نمی‌باشند. الگوریتم رگرسیون تنها زمانی که تعداد ماتریس‌های پایه کمتر از اندازه پرس‌وجوی کاربر باشد، جواب دارد. الگوریتم min-cut نیز تنها جستجوهای مربوط به یک انجمن را پاسخ می‌دهد، به عبارت دیگر min-cut در حالتی که همه‌ی برچسب‌های بردار Y یکسان باشند، به جواب می‌رسد ($Q=[1]_{n\times n}$).

مهمترین مزیت الگوریتم ما در استخراج رابطه، قابلیت تفسیر آن می‌باشد. یک مجموعه داده مانند DBLP که [13] در شبیه‌سازی از آن استفاده کرده است، بیش از ۲۲۰۰ رابطه پایه دارد. در صورتی که به طور معمول از رگرسیون استفاده شود، ۲۲۰۰ ضریب برای رابطه هدف بدست می‌آید و بسیاری از این روابط، در رابطه هدف تاثیر گذارند، که کسرت این روابط باعث عدم قابلیت تفسیر این روابط می‌شود. حال اگر کسی بخواهد تعداد ماتریس‌های‌پایه که در یک راه‌حل استفاده می‌شود را به K ماتریس محدود کند، مساله تبدیل به یافتن، K ماتریس پایه‌ای می‌شود که ترکیب خطی آنها بیشترین شباهت به ماتریس هدف را دارد. راه‌حل این مساله همان راه‌حلی است که در قسمت ۳ ارائه شد. انگیزه این کار، این است که یک فرد ممکن است، بخواهد K تا ماتریس پایه‌ای که بیشترین تاثیر روی ماتریس هدف را دارد، پیدا کند. در نگاه اول به این مساله دو راه‌حل بدیهی به نظر می‌رسد. راه‌حل اول این است که هر Kتایی از این n تا ماتریس پایه انتخاب شود و در نهایت بهترین جواب انتخاب شود. واضح است که این راه‌حل $\binom{n}{k}$ بار نیاز به اجرا دارد و این زمان اجرای زیادی است. راه‌حل دیگر، حل مساله با n تا ماتریس پایه و انتخاب K تا، بزرگترین ضریب می‌باشد. مثال زیر نشان می‌دهد که در این حالت، لزوماً بهترین راه‌حل ایجاد نمی‌شود.

حالت ساده‌ای را فرض کنید که ما سه رابطه پایه داریم، ماتریس هدف (Q) و ماتریس‌های پایه($M_i$) در شکل ۱ نشان داده شده‌اند. نتیجه استفاده از رگرسیون برای حل این مساله عبارت است از: $Q=M_2+M_3$. حال فرض کنید که K=1 است و ما می‌خواهیم ماتریس هدف، حداکثر توسط یک ماتریس پایه بیان شود. با مرتب کردن ضرایب نتیجه و انتخاب بزرگترین ضریب، حاصل را برابر $Q=M_2$ یا $Q=M_3$ بدست می‌آوریم، در حالی که بهترین نتیجه در حالت K=1، ماتریس $Q= M_1$ است.

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.
www.SID.ir

$M_1$=

| | | | | |
|---|---|---|---|---|
| * | 1 | 1 | 0 | 0 |
| 1 | * | 1 | 0 | 0 |
| 1 | 1 | * | 0 | 0 |
| 0 | 0 | 0 | * | 1 |
| 0 | 0 | 0 | 0.9 | * |

Q=

| | | | | |
|---|---|---|---|---|
| * | 1 | 1 | 0 | 0 |
| 1 | * | 1 | 0 | 0 |
| 1 | 1 | * | 0 | 0 |
| 0 | 0 | 0 | * | 1 |
| 0 | 0 | 0 | 1 | * |

$M_3$=

| | | | | |
|---|---|---|---|---|
| * | 0 | 0 | 0 | 0 |
| 0 | * | 1 | 0 | 0 |
| 0 | 1 | * | 0 | 0 |
| 0 | 0 | 0 | * | 1 |
| 0 | 0 | 0 | 1 | * |

$M_2$=

| | | | | |
|---|---|---|---|---|
| * | 1 | 1 | 0 | 0 |
| 1 | * | 0 | 0 | 0 |
| 1 | 0 | * | 0 | 0 |
| 0 | 0 | 0 | * | 0 |
| 0 | 0 | 0 | 0 | * |

**شکل (۱): مثال عدم کارایی رگرسیون**

پیچیدگی زمانی رگرسیون عبارتست از:

$$T= n^3 + 2\ n^2k + 2nk = O(n^3+ n^2k) \quad (۳)$$

که n، تعداد ماتریس‌های پایه و k اندازه پرس‌وجوی کاربر است. برای الگوریتم ژنتیک، اگر G، تعداد نسل‌ها، P اندازه جمعیت و $t_1$ و $t_2$ به ترتیب احتمال mutation و crossover را بیان کنند و n هم تعداد ماتریس‌های پایه باشد، زمان اجرای الگوریتم ما:

$$G*P*t_1*1 + G*P*t_2*n = O(n) \quad (۴)$$

خواهد بود. زمان مورد نیاز برای انجام mutation، O(1) و برای crossover برابر O(n) است.

## ۶- نتایج تجربی

برای بدست آوردن نتایج خروجی از مجموعه داده FilmTrust استفاده کرده‌ایم[27]. این مجموعه داده یک وب‌سایت است که افراد در مورد فیلم‌های مختلف نظر داده و به نظرات دیگران مقداری از اعتماد بین یک تا ده، تخصیص می‌دهند. در مورد هر فیلم نیز مقدار اعتمادی بین صفر تا چهار می‌دهند. این مجموعه داده حدود ۵۶۰ کاربر عضو دارد.

در این مجموعه داده، نظرات در مورد هر فیلم به عنوان یک رابطه (شبکه اجتماعی)، افراد درگیر در آن به عنوان گره‌های شبکه و میزان اعتماد افراد به آن فیلم به عنوان وزن یالها در نظر گرفته می‌شود. از آنجایی که باید خروجی روش پیشنهادی ما، با یک مقدار واقعی مقایسه شود، نیاز به یک معیار مقایسه داریم. برای این کار، در هر بار اجرای الگوریتم یک رابطه را از ماتریس‌های پایه حذف کرده و به عنوان پرس‌وجو به الگوریتم می‌دهیم، اختلاف خروجی الگوریتم با مقدار واقعی موجود بر یالهای ماتریس پرس‌وجو، میزان خطای الگوریتم ما را مشخص می‌کند که آن را $\Delta$ می نامیم. اگر میانگین $\Delta$ برای تمامی ماتریس‌های پایه را $\Delta'$ بنامیم، خروجی حاصل از روش پیشنهادی ما که در قسمت ۳-۵ آورده شده است و نیز اجرای الگوریتم Tidal بدون استفاده از استخراج رابطه در جدول ۱ آمده است.

**جدول (۱):خروجی اجرای الگوریتم‌ها**

| الگوریتم | $\Delta'$ |
|---|---|
| الگوریتم ۱ | ۱/۶۳ |
| الگوریتم ۲ | ۱/۲۷ |
| Tidal | ۲/۴۳ |

همانطور که مشاهده می‌کنید، میزان خطای تخمین اعتماد بین گره‌ها در الگوریتم ۲ از روشهای دیگر بهتر می‌باشد. همچنین نتیجه استخراج رابطه در الگوریتم۱ و ۲ و سپس استنتاج اعتماد بهتر از روش Tidal می‌باشند. علت اصلی این امر این است که استخراج رابطه، باعث استخراج انجمن مخفی[vii] می‌شود که در بین گره‌ها وجود دارد.

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.
www.SID.ir

## ۷- نتیجه

ما ابتدا با به کارگیری دو الگوریتم ژنتیک، به استخراج رابطه و تشکیل رابطه هدف پرداختیم. با انجام این کار، علاوه بر اینکه بسیاری از رابطه‌هایی را که نسبت به روابط دیگر مهمتر بودند، در رابطه هدف وارد نمودیم. از طرف دیگر، پرس‌وجوی کاربر را نیز در ایجاد رابطه هدف، در نظر گرفتیم و با توجه به پرس‌وجوی کاربر یک رابطه هدفمند را از روی روابط پایه با استفاده از وزن‌دادن به هر رابطه پایه ساختیم. در انتها نیز، ادعای خود را با استفاده از یک مجموعه داده واقعی ثابت کردیم. الگوریتم پیشنهادی ما، نسبت به روش‌های مرسوم دیگر، علاوه بر اینکه در عمل نتایج بهتری را بدست آورد، دارای زمان اجرای مناسب، O(n) می‌باشد.

در این مقاله با تمامی ماتریس‌های پایه، هنگام استخراج رابطه به طور یکسان برخورد می‌شود؛ به عنوان یک کار آتی در صورتی که بر اساس برخی خواص ماتریس‌ها، بین آنها تمایز قائل شویم، نتایج بهتری می‌توان بدست آورد. به عنوان مثال، ماتریس‌هایی که تمامی درایه‌های آن یک است، اطلاعات کمتری نسبت به ماتریسی که تقریبا درایه‌های صفر و یک آن برابر است، دارد.

## تشکر و قدردانی

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.

## مراجع

[1] Karlins, M. & Abelson, H. I., *Persuasion, how opinion and attitudes are changed*, Crosby Lockwood & Son, 1970.
[2] Buskens, V., *The Social Structure of Trust*, Social Networks, 1998.
[3] Marimon, R., Nicolini, J. P. & Teles, P, *Competition and Reputation*, Proceedings of the World Conference Econometric Society, Seattle, 2000.
[4] Gambetta, D., *Can We Trust Trust?* Trust: Making and Breaking Cooperative Relations, pages 213.238. Basil Blackwell. Oxford, 1990.
[5] Mui, L., Mohtashemi, M., and Halberstadt, A., *A computational model of trust and reputation*, In Proceedings of the 35th International Conference on System Science, pages 280–287, 2002.
[6] Pujol, J., Sanguesa, R., Delgado, J., *Extracting reputation in multiagent systems by means of social network topology*, In Proceedings of First InternationalJoint pages 467-474, 2002.
[7] Sabater, J., Sierra, C., *Reputation and social network analysis in multi-agent systems*, Proceedings of the first international joint conference on Autonomous agents and multiagent systems, 2002.
[8] Sabater, J., Sierra, C., *Social ReGreT, a reputation model based on social relations*, ACM SIGecom Exchanges, vol. 3.1, pp. 44-56, 2002.
[9] Mui, L., *Computational Models of Trust and Reputation Agents, Evolutionary Games, and Social Network*, BS thesis MIT University, 2003.
[10] Milgram, S., *The small world problem*, Psychology Today 2, 60–67, 1967.
[11] Newman, M., *Models of the small world*, J. Stat. Phys. 101, 819–841, 2000.
[12] Lada, A., *The small world web*, In Proc. of ECDL, pages 443–452, September, 1999.
[13] Cai, D., Shao, Z., He, X., Yan, X., Han, J., *Community Mining from Multi-Relational Networks*, Principles and Practice of Knowledge Discovery in Databases (PKDD '05), Porto, Portugal, Oct., 2005.
[14] Newman, M. E. J., *Finding community structure in networks using the eigenvectors of matrices*, Physical Review E, vol. 74, Issue 3, id. 036104, 2006.
[15] Boguna, M., Pastor-Satorras, R., Diaz-Guilera, A., Arenas, A., *Emergence of clustering, correlations, and communities in a social network model.* Preprint condmat/0309263, 2003.
[16] Tyler, J. R., Wilkinson, D. M., Huberman, B. A., *Automated discovery of community structure within organizations.* In Proceedings of the First International Conference on Communities and Technologies, 2003.
[17] Garey, M. R., Johnson, D. S., *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, San Francisco, 1979.
[18] Scott, J., *Social Network Analysis*, A Handbook. Sage Publications, London, 2nd edition , 2000.
[19] McKnight, D.H., Chervany, N.L., *The Meanings of Trust*. Technical Report MISRC Working Paper Series 96-04, University of Minnesota, Management Information Systems Reseach Center, 1996.
[20] Mui, L., Mohtashemi, M., and Halberstadt, A., *A computational model of trust and reputation,* In Proceedings of the 35th International Conference on System Science, pages 280–287, 2002.
[21] Marsh, S., *Formalising Trust as a Computational Concept*, PhD thesis, University of Stirling, 1994.
[22] Abdul-Rahman, A., Hailes, S., *Supporting trust in virtual communities*, In: Proc. of Hawaii International Conference on System Sciences, 2000.

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.
www.SID.ir

[23] Sabater, J., Sierra, C., *Review on computational trust and reputation models*, Artificial Intelligence Review 24(1) 33–60, 2005.
[24] Sabater, J., Sierra, C., *Reputation and social network analysis in multi-agent systems*, In: Proc. of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems, New York, NY, USA, ACM Press 475–482, 2002.
[25] Sabater, J., *Trust and reputation for agent societies*, PhD thesis, Institutd'Investigacion en Intelligencia Artificial, Spain, 2003.
[26] Golbeck, J., *Computing and Applying Trust in Web-Based Social Networks*, PhD thesis, University of Maryland, College Park, 2005.
[27] mindswap. *Film Trust*, ptc/08-08-07, August 2007, http://trust.mindswap.org/FilmTrust.

[i] Trust
[ii] Reputation
[iii] Relation extraction
[iv] Target relation
[v] Agent
[vi] Community
[vii] Hidden community

این تحقیق با حمایت مرکز تحقیقات مخابرات ایران انجام پذیرفته است.
www.SID.ir