فصل دهم

موضوع : Mining Social-Network Graphs

به نام خدا

مقدار زیادی داده با تجزیه و تحلیل از شبکه های اجتماعی به دست می آید.

شناخته شده ترین نمونه یک شبکه اجتماعی ، رابطه "دوستان" است که در سایت هایی مانند فیس بوک یافت می شود.

با این حال ، همانطور که خواهیم دید بسیاری از منابع داده دیگر وجود دارند که افراد یا نهادهای دیگر را به هم وصل می کنند.

در این فصل تکنیک های تحلیل داده های موجود در شبکه ها را بررسی می کنیم.

یک سوال مهم در مورد یک شبکه اجتماعی ، چگونگی شناسایی "اجتماعات" ، یعنی زیر مجموعه گره ها (افراد یا اشخاص دیگری که شبکه را تشکیل می دهند) با اتصالات غیرمعمول قوی است.

برخی از تکنیک های مورد استفاده برای شناسایی جوامع مشابه الگوریتم های خوشه بندی است که در فصل 7 درباره آنها گفتیم.

با این حال ، مجموعه ها تقریباً هرگز قسمت هایی از گره ها را در یک شبكه تقسیم نمی كنند.

در عوض ، مجموعه ها معمولاً با هم همپوشانی دارند.

به عنوان مثال ، شما ممکن است به چندین گروه مختلف از دوستان یا همکلاسی های خود تعلق داشته باشید.

مردم یک جامعه تمایل دارند یکدیگر را بشناسند ، اما مردم دو جامعه مختلف به ندرت یکدیگر را می شناسند.

شما نمی خواهید فقط در یکی از گروه ها قرار بگیرید و همچنین معقول نیست که همه ی افراد از همه ی جوامع خود را حتما در یک گروه قرار دهید.

همچنین در این فصل الگوریتم های کارآمد را برای کشف سایر خصوصیات گراف ها بررسی می کنیم.

ما به یک مفهومی تحت عنوان simrank نگاه می کنیم که این مفهوم به معنی کشف شباهت بین گره های یک گراف است.

ما شمارش مثلثی را به عنوان روشی برای سنجش ارتباط در یک جامعه استفاده می کنیم.

ما برای اندازه گیری دقیق و تقریبی اندازه مجموعه ها و گره های در یک گراف ، الگوریتم های مناسبی را ارائه می دهیم.

سرانجام ، برای محاسبه بستار انتقالی ، به الگوریتم های ویژه ای را بررسی می کنیم.

10.1 شبکه های اجتماعی به صورت گراف

بحث خود را در مورد شبکه های اجتماعی با معرفی یک مدل گرافی آغاز می کنیم.

هر نمودار گرافی مناسب به نمایش یک اجتماع در شبکه های اجتماعی نیست.

بنابراین، ما درباره ی ایده ی اصل محلیت که جز ویژگی های اصلی شبکه های اجتماعی است با کمک نودها و یال ها در گراف ها صحبت می کنیم. با کمک گره ها و یال ها تمایل خوشه بندی در شبکه ها بررسی می کنیم.

در این بخش همچنین برخی از انواع مختلف شبکه های اجتماعی که در عمل مورد استفاده قرار می گیرند را بررسی می کنیم.

10.1.1 شبکه اجتماعی چیست؟

وقتی به یک مفهوم شبکه اجتماعی فکر می کنیم ، به فیس بوک ، توییتر ، Google+ یا وب سایت دیگری فکر می کنیم که "شبکه اجتماعی" نامیده می شود و در واقع این شبکه ها نماینده ای از مفهوم شبکه های اجتماعی هستند.

ویژگی های اساسی یک شبکه اجتماعی عبارتند از:

1. مجموعه ای از اشخاص در شبکه های اجتماعی در شبکه وجود دارند که به طور معمول این موجودات مردم هستند اما می توانند چیزهای دیگری نیز باشند؛ در بخش 10.1.3 به مثالهای بیشتری در این مورد می پردازیم.
2. حداقل یک رابطه بین موجودیت های موجود (کاربران) در شبکه های اجتماعی وچود دارد. به رابطه ی بین موجودیت ها در فیس بوک Relationship می گویند. ارتباط ها یا وجود داشته یا ندارد پس دو نفر یا دوست هستند یا نیستند؛  
   گرچه در گراف های شبکه های اجتماعی رابطه ها دارای یک درجه هستند.  
   این درجه ها می تواند مقدار گسسته داشته باشد. به عنوان مثال در گوگل پلاس این درجه ها با عنوان های دوستان، خانواده، آشنایان و ... شناخته می شود.  
   این درجه می تواند یک عدد صحیح باشد یا یک عدد کسری باشد که از میزان صحبت بین دونفر بدست می آید.
3. محلیت در این شبکه ها به صورت غیرتصادفی است. این شرط برای نرمال سازی سخت ترین شرط است. اما ارتباطش به گرایش خوشه ها و روابط آنها می پردازد. یعنی اگر موجودیت A به هر دو موجودیت B و C مربوط باشد احتمال رخداد آن از میانگین B و C بدست می آید.

10.1.2 شبکه های اجتماعی به صورت گراف

شبکه های اجتماعی به طور معمول به عنوان گراف هایی مدل سازی می شوند که بعضا از آنها به عنوان یک گراف اجتماعی یاد می کنیم.

در این نمودار گراف گره ها موجودیت ها هستند و یال ها اتصال و ارتباط بین دو موجودیت (کاربر) هستند. اگر در این گراف درجه ای موجود باشد روی یال ها برچسب زده می شود.

غالبا گراف های موجود در شبکه های اجتماعی بدون جهت هستند؛ مثل گراف دوستان در شبکه اجتماعی فیس بوک.

اما می توان گراف های جهت دار هم داشت؛ مانند نمودار فالورهای توییتر یا گوگل پلاس.

مثال 10.1 : شکل 10.1 نمونه ای از یک گراف کوچک شبکه اجتماعی است؛ گره های این گراف از A تا G نامگذاری شده است. رابطه ی موجود در این گراف به عنوان دوست شناخته می شود که یال ها را تشکیل می دهد.

به عنوان مثال موجودیت B با A و C و D دوست می باشد.

آیا این گراف واقعا نشان دهنده ی یک شبکه ی اجتماعی و نمایش روابط آنها است؟

ابتدا توجه داشته باشید که گراف زیر دارای 9 یال می باشد.



شکل 10.1 : نمونه ای از شبکه اجتماعی کوچک

. در واقع 21 جفت گره می تواند در این شبکه دارای لبه باشد یا حداکثر یال های موجود در این شبکه 21 باشد.

فرض کنید X و Y و Z نودهای شکل 10.1 هستند که بین x و y و همچنین بین x و z دارای یال می باشد.

احتمال وجود یال بین y و Z چقدر است؟

اگر این نمودار بزرگ باشد این احتمالا به صورت کسری می باشد. یعنی در واقع 9 تقسیم بر 21 که برابر 0.429 می شود.

اما ، از آنجا که نمودار كوچك است ، بین احتمال واقعی و نسبت تعداد لبه ها به تعداد جفت گره ها تفاوت قابل ملاحظه ای وجود دارد.

از آنجایی که می دانیم یال های (X، Y) و (X، Z) وجود دارد ، تنها 7 یال باقی مانده است. این 7 یال می توانند بین هر 19 گره باقی مانده از گره ها قرار بگیرند.

بنابراین احتمال یک لبه بین z و y برابر 7 تقسیم بر 19 که برابر 0.368 می شود است.

حال باید این احتمال را محاسبه کنیم که لبه (Y ، Z) در شکل 10.1 وجود داشته باشد ، با توجه به اینکه لبه ها (X ، Y) و (X ، Z) وجود دارند. آنچه در واقع باید حساب کنیم ، جفت گره هایی است که می توانند Y و Z باشند ، بدون اینکه نگرانی در مورد کدام گره Y باشد و کدام Z باشد. اگر X A باشد ، باید Y و Z به ترتیب B و C باشند. از آنجا که لبه (B ، C) وجود دارد ، A یک مثال مثبت (که در آن لبه وجود دارد) است و هیچ نمونه منفی (جایی که لبه وجود ندارد) در آن نیست.

در مواردی که X به جای C ، E یا G است نتایج یکسان است.

در هر حالت ، X فقط دو همسایه دارد و لبه بین همسایگان وجود دارد. بنابراین ، ما تاکنون چهار نمونه مثبت و صفر مثال منفی را دیده ایم.

حال ، در نظر بگیرید که X = F. F دارای سه همسایه ، D ، E و G است. لبه هایی بین دو سه جفت همسایه وجود دارد ، اما هیچ لبه ای بین G و E وجود ندارد. بنابراین ، ما دو نمونه مثبت دیگر را می بینیم و اولین نمونه منفی خود را می بینیم.

اگر X = B ، دوباره سه همسایه وجود دارد ، اما فقط یک جفت همسایه ، A و C ، یک لبه دارند. بنابراین ، ما دو مثال منفی دیگر ، و یک مثال مثبت ، برای کل هفت مثبت و سه منفی داریم.سرانجام ، وقتی X = D ، چهار همسایه وجود دارد. از شش جفت همسایه ، فقط دو نفر بین آنها لبه دارند.

بنابراین ، تعداد کل نمونه های مثبت 9 و تعداد کل نمونه های منفی 7 است. در شکل 10.1 می بینیم که کسر ما برابر 9 تقسیم بر 16 است که در واقع برابر 0.563 می شود. این کسر خیلی بیشتر از مقدار قابل انتظار ما که 0.368 است می باشد.

در نتیجه شکل 10.1 واقعا اصل locality در شبکه های اجتماعی را نشان می دهد.