بسمه تعالی

پژوهشکده آماد، فناوری دفاعی و پدافند غیرعامل

گروه علوم و فناوری‌های دفاعی

میز تخصصی رصد و پایش علم، فناوری و نوآوری دفاعی و امنیتی

پیش‌بینی پیوند موضوعی پژوهش‌های علمی با استفاده از تحلیل شبکه‌ی مقالات علمی

رضا شکرچیان چالشتری
E-mail: rz.shekarchian@ut.ac.ir

گزارش ماهانه شماره (۳) – ۹۷/۷/۲۰

چکیده

شبکه‌های پیچیده مدل ریاضی شبکه‌های واقعی می‌باشند که روش‌ها و الگوریتم‌های مختلفی برای تحلیل این شبکه‌ها قابل استفاده می‌باشند. شبکه‌ی پیچیده شامل چندین قسمت مجزا می‌باشد که به صورت غیرخطی با یکدیگر در ارتباط می‌باشند. مغز انسان شبکه‌ای از سلول‌های عصبی می‌باشد، اجتماع مجموعه‌ای از انسان‌ها می‌باشد که با رابطه‌هایی مانند رابطه‌ی دوستی، همکاری و ... به یکدیگر متصل می‌شوند. دراینترنت صفحات وب با استفاده از هایپرلینک‌ها شبکه‌ی وب را می‌سازند. تئوری‌های گراف از جمله ابزارهای اساسی در تحلیل شبکه‌های پیچیده می‌باشند. در ادامه تعاریفی از گراف آورده می‌شود. انواع مختلف گراف که برای نمایش شبکه‌ی پیچیده مورد استفاده قرار می‌گیرند معرفی می‌شوند. برخی از ویژگی‌هایی که در بسیاری از شبکه‌های پیچیده مشترک می‌باشند بررسی می‌شوند. برخی از مدل‌سازی‌های این شبکه‌ها معرفی می‌شوند و در ادامه به برخی از عملیات تحلیلی بر روی شبکه‌های پیچیده اشاره می‌شود. در این تحقیق عملیات تحلیلی‌ای که استفاده می‌شود پیش‌بینی لینک می‌باشد.

کلمات کلیدی: گراف، شبکه‌های پیچیده، تحلیل شبکه

۱. شبکه‌های پیچیده

در [1] یک شبکه را مجموعه‌ای از نودها و یال‌ها تعریف کرده است. نودها نشان دهنده‌ی موجودیت‌ها و بخش‌های یک سیستم و یال‌ها نمایان‌گر ارتباط میان نودها می‌باشند. به این مجموعه گراف نیز گفته می‌شود. سیستم‌های بسیاری وجود دارند که می‌توان آن‌ها را در قالب گراف نمایش داد که شناخته شده‌ترین آن‌ها دنیای وب می‌باشد. از جمله شبکه‌های دیگر می‌توان به شبکه‌ی غذایی، شبکه‌ی ارتباط پروتئین‌ها، ارتباط سیستم‌های عصبی و ... اشاره کرد. از شبکه‌های شناخته شده شبکه‌های اجتماعی می‌باشند که که لینک‌ها در آن‌ها نشان دهنده‌ی ارتباط دوستی، همکاری، مالی و ... می‌باشد. نوع دیگر شبکه، گراف حاصل از نظرات مردم بر روی فیلم‌ها، بازیگران، محصولات و ... می‌باشد.

دسته‌بندی‌های مختلفی برای گراف‌ها بر اساس نوع لینک‌ها، جهت‌دار بودن لینک‌ها، وزن‌دار بودن و تعداد لینک‌ها و ... وجود دارد.

۱- گراف‌های ساده: یال‌ها بدون جهت و بدون وزن می‌باشند. این یال‌ها متقارن می‌باشند. ارتباط میان نودها صفر و یکی می‌باشد.

۲- گراف جهت‌دار: یال میان نودها جهت‌دار می‌باشد.

۳- شبه گراف: گراف‌هایی هستند که احتمال وجود چندین یال بین دو نود در آن‌ها وجود دارد به این گراف‌های چند-گراف[۱] هم گفته می‌شود. همچنین امکان وجود یال از یک نود به همان نود نیز وجود دارد. می‌توانند هم یال جهت‌دار و هم یال بدون جهت داشته باشند. یک چند-گراف را می‌توان با چند لایه گراف نشان داد که در هر لایه گراف‌ها نودهای مشابه و فقط یک نوع لینک دارند. این گراف‌ها به گراف‌های چندگانه[۲] نیز شناخته می‌شوند.

۴- گراف وزن‌دار[2, 3]**:** به یال‌ها وزن نسبت داده می‌شود. در بعضی شرایط می‌توان چند-گراف را به گراف وزن‌دار تبدیل کرد. مثلا تعداد یال‌هایی که دو نود را به یکدیگر وصل می‌کند می‌تواند به عنوان وزن در گراف وزن‌دار محسوب شود.

۵- فراگراف[۳]: گراف‌هایی هستند که یک یال بیش از دو نود را به یکدیگر متصل می‌کند. به این یال‌ها فرالینک[۴] گفته می‌شود. گرافی را در نظر بگیرید که کاربران، منابع و برچسب‌ها نودهای آن باشند. برچسب‌ها کلماتی هستند که توسط کاربر بر روی منابعی که در شبکه استفاده می‌کنند، گذاشته می‌شوند. یک یال در این شبکه ترکیبی از یک نود کاربر، یک نود منبع و یک یک نود برچسب می‌باشد و رابطه‌ای میان سه موجودیت را نشان می‌دهد.

جزئیات بیشتر راجع به ساختار گراف‌ها و کاربردهای آن‌ها در [1] قابل دست‌یابی می‌باشد.

یکی از مثال‌های گراف ساده شبکه‌ی همکاری می‌باشد. نودها نمایانگر نویسندگان می‌باشد. بین دو نویسنده یال وجود دارد اگر حداقل یک مقاله‌ی مشترک داشته باشند. همچنین می‌توان از گراف وزن‌دار برای مدل کردن ارتباط نویسندگان استفاده کرد. وزن یال‌ها تعداد مقاله مشترک میان دو نویسنده می‌باشد. همچنین می‌توان ارتباط میان نویسندگان را با چند-گراف مدل کرد. بین دو نویسنده چندین یال امکان دارد وجود داشته باشد که نشان‌گر مقاله‌های مشترک میان آن‌ها می‌باشد. هر یال می‌تواند ویژگی‌هایی همچون زمان مقاله یا ژورنالی که در آن چاپ شده را داشته در خود داشته باشد.

نوع دیگری از دسته‌بندی شبکه‌ها می‌تواند بر اساس نوع ارتباط نودها باشد. مجموعه‌ای از نودها ساخته می‌شود که یک نود هیچ‌وقت به نودی از مجموعه‌ی خود متصل نمی‌شود. ولی می‌تواند به نودهای دیگر در دیگر دسته‌ها متصل شود. این مفهوم این امکان را فراهم می‌کند که دسته‌بندی جدیدی از گراف‌ها داشته باشیم:

---

۱ multi-graph
۲ multiplex-graph
۳ hypergraph
۴ hyperlink

۱- گراف یکپارچه: این گراف فقط یک مجموعه نود دارد و هیچ دسته‌بندی‌ای از نودها ندارد. این امکان برای هر نود وجود دارد که به نود دیگر یال داشته باشد. شبکه‌ی همکاری نویسندگان مقالات علمی مثالی از این نوع گراف می‌باشد.

۲- گراف دوبخشی: گراف دوبخشی دو مجموعه از نودها دارد و یک نود از یک مجموعه تنها می‌تواند به نودی از مجموعه‌ی دیگر لینک داشته باشد. میان لینک‌های یک مجموعه هیچ لینکی وجود ندارد. مثالی از گراف دوبخشی گراف مقاله-نویسنده باشد. دو مجموعه از نودها وجود دارد. نودهای یک مجموعه نمایانگر مقالات و نودهای مجموعه‌ی دیگر نمایانگر نویسندگان می‌باشد. یال‌ها فقط می‌توانند میان نویسندگان و مقالات باشند که نشان می‌دهد یک نویسنده در نوشتن کدام مقاله نقش داشته است. مثال دیگر گراف کاربر-کالا می‌باشد که برای تحلیل بازار به کار برده می‌شود. نودها در دو دسته‌ی کاربران و کالاها قرار می‌گیرند. یک کاربر به یک کالا یال دارد اگر آن را خریده باشد. تحلیل گراف در این‌جا می‌تواند برای توصیه‌ی کالا به کاربر بر اساس انتخاب‌های کابر استفاده شود . این امکان وجود دارد که دو گراف یکپارچه از گراف دوبخشی ساخت. لینک‌ها در هر گراف یکپارچه بر اساس لینک‌های گراف دوبخشی ساخته می‌شوند. به طور مثال گراف مقاله-نویسندگان می‌تواند دو گراف یکپارچه داشته باشد. یکی فقط ساخته شده از نودهای نویسندگان و دیگری ساخته شده از نودهای مقالات. در گراف نویسندگان بین دو نویسنده یال وجود دارد، اگر در گراف دوبخشی به حداقل یک مقاله یال مشترک داشته باشند. به طور مشابه در گراف مقالات بین دو مقاله لینک وجود دارد اگر در گراف دوبخشی حداقل به یک نویسنده‌ی مشترک یال داشته باشند.

۳- گراف سه‌بخشی: گراف سه‌بخشی سه مجموعه نود دارد. مثال کاربر، محصول، برچسب را می‌توان با گراف سه‌بخشی نمایش داد.

گراف‌هایی که بیشتر از ۳ مجموعه نود دارند، گراف چندبخشی نامیده می‌شوند. شکل ۱ انواع گراف را نمایش می‌دهد.

۱-۱ تعاریف ریاضی

یک گراف ساده به صورت $G=< V, E >$ نمایش داده می‌شود که $V = v_1 , v_2 , ..., v_n , |V| = N$ مجموعه‌ای محدود از نودها می‌باشد. $E \subseteq V \times V, E = (v_i , v_j), i \neq j, |E| = M$ مجموعه‌ای از یال‌ها در گراف می‌باشد. گراف G را می‌توان با ماتریس مجاورت NxN نمایش داد. در این ماتریس درایه‌ها بر اساس وجود یا عدم وجود لینک بین دو نود می‌توانند ۱ یا ۰ باشند.

گراف وزن‌دار را می‌توان به صورت $G =< V, E, W >$ نمایش داد. پارامتر $W = w : E \rightarrow R$ که w تابعی است که مقداری به عنوان وزن به یال نسبت می‌دهد.

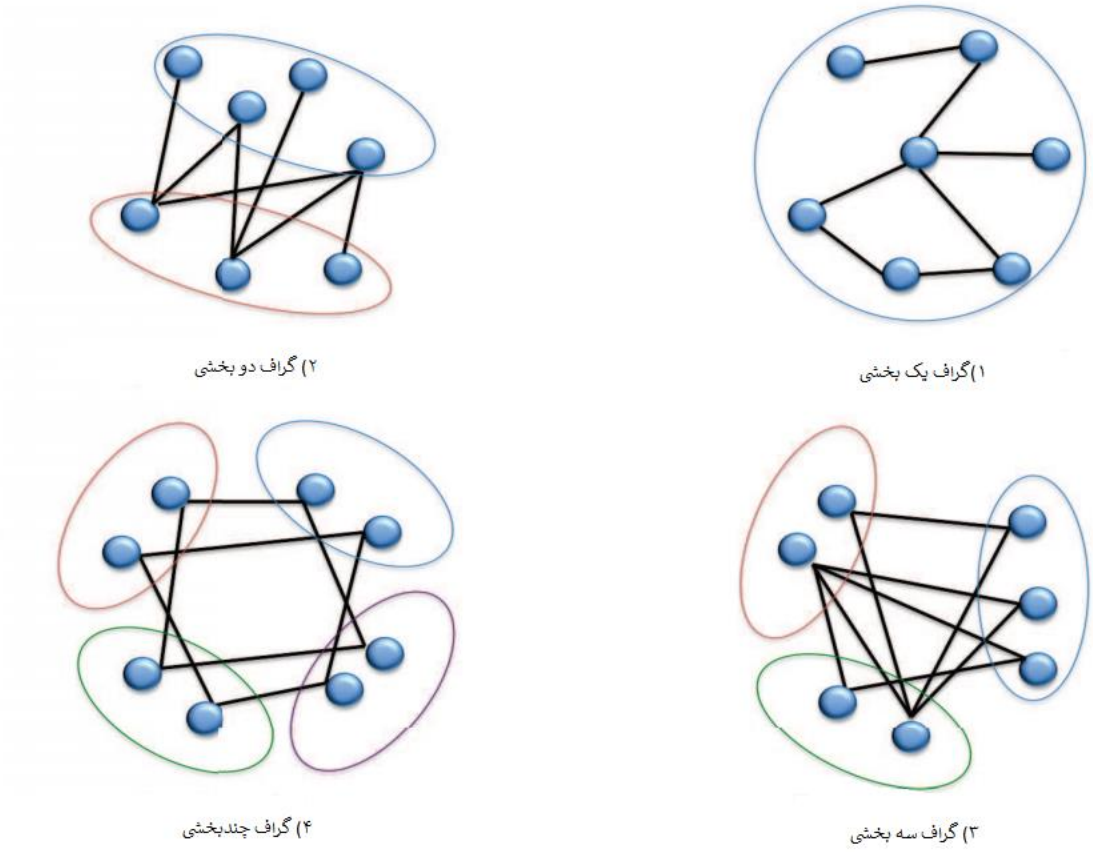
در گراف بدون جهت برای یک یال $(v_i , v_j) \Leftrightarrow (v_j , v_i)$ می‌باشد و ماتریس مجاورت متقارن می‌باشد. در گراف جهت‌دار هر یال جهتی دارد. همسایگان یک نود در گراف مجموعه‌ای از نودها می‌باشند که مستقیما به نود متصل شده‌اند. مجموعه همسایگان نود v_i به صورت

$$\Gamma(vi) = \{vj : (vi, vj) \in E\} \qquad ۱-۱$$

در گراف جهت‌دار درجه‌ی یک نود به دو بخش *درجه ورودی* و *درجه خروجی* بر اساس جهت یال تقسیم می‌شود.

یک مسیر بین دو نود v_0 و v_k در یک گراف ساده گرافی به صورت $P=(V_p , E_p)$ می‌باشد. مجموعه‌ای از نودهای $V_p = v_0 , v_1 , v_2 , ..., v_{k-1} , v_k$ و یال‌های $E_p=(v_0 , v_1), (v_1 , v_2), ..., (v_{k-1} , v_k)$ که $V_p \subseteq V$ و $E_p \subseteq E$. در اینجا V_p در اصل مجموعه‌ای از نودها می‌باشد که در آن نود V_i به نودهای قبلی و بعدی مستقیما متصل است. طول یک مسیر تعداد یال‌ها در E_p می‌باشد. کوتاه‌ترین مسیر بین دو نود، مسیری با کمترین طول می‌باشد و فاصله‌ی دو نود را مشخص می‌کند.

گراف G پیوسته[۱] نامیده می‌شود اگر برای هر دو نود $v_i , v_j \in V$ مسیری از v_i به v_j وجود داشته باشد. گراف‌ها معمولا پیوسته نیستند. ولی می‌توان آن‌ها را ترکیبی از چند زیرگراف پیوسته در نظر گرفت. این زیرگراف‌ها مجموعه‌های پیوسته[۲] نامیده می‌شوند.



شکل ۱) انواع مختلف گراف

[۱] connected

[۲] connected components

معنای علامت	علامت‌ها
گراف	G
مجموعه نودها	V
مجموعه یال‌ها	E
تعداد نودها	N
تعداد یال‌ها	M
مجموعه‌ی مسیرها بین نود V_I و نود V_J	$Paths(v_i , v_j)$
مجموعه نودها در مسیر P	V_p
مجموعه یال‌ها در مسیر P	E_p
درجه نود V_I	K_i
طول کوتاه‌ترین مسیر بین نودر V_I و نود V_J	$Dist(v_i , v_j)$

جدول ۱) علائم

۲. ویژگی‌های شبکه‌های پیچیده:

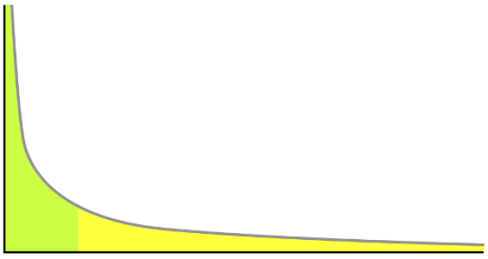
بیشتر انواع شبکه‌های پیچیده ویژگی‌های توپولوژی مشترکی دارند. در ادامه به چند مورد آن‌ها اشاره می‌شود:

۱- همبستگی[۱]:نودها در شبکه‌های پیچیده معمولا به خوشه‌هایی از مجموعه‌های پیوسته تقسیم می‌شوند. در مجموعه‌ی پیوسته تمام نودها به صورت مستقیم یا غیر مستقیم به یکدیگر متصل می‌باشند. در یک گراف معمولا یک یا دو مجموعه‌ی پیوسته بسیار بزرگ و تعداد زیادی مجموعه‌ی پیوسته‌ی کوچک وجود دارد.

۲-درجه توزیع: درجه توزیع احتمال این است که یک نود k همسایه در شبکه داشته باشد. به عبارت دیگر احتمال اینکه درجه‌ی یک نود k باشد. معمولا شبکه‌های پیشرفته درجه توزیعی دارند که از قانون قدرت[۲] پیروی می‌کند. شکل ۲ شکلی کلی از نمودار قانون قدرت می‌باشد. همان‌طور که دیده می‌شود در این توزیع تعداد زیادی از نودها درجه‌ی پایینی دارند و تعداد کمی از نودها درجه‌ی بسیار بالایی دارند. ضریب قانون قدرت مشخص کننده که نرخ کاهش درجه در نمودار می‌باشد. هرچه ضریب بالاتر باشد، احتمال پیدا کردن نودی با درجه‌ی بالا کمتر می‌شود.

۱ Connectedness

۲ power law



شکل ۲) توزیع قانون قدرت https://en.wikipedia.org/wiki/Power_law#/media/File:Long_tail.svg

۳- ضریب خوشه‌بندی: در بسیاری از شبکه‌ها دیده می‌شود که دو نودی که به یک نود مشترک اتصال می‌باشند، تمایل دارند بین یکدیگر لینک برقرار کنند. به این ویژگی انتقال‌پذیری گفته می‌شود و توسط ضریب خوشه‌بندی اندازه‌گیری می‌شود. ضریب خوشه‌بندی اندازه‌گیری می‌کند که همسایگان یک نود با چه احتمالی به یکدیگر متصل می‌شوند. طبق تعریفی که در [4] آمده ضریب خوشه‌بندی نود $v_i \in V$ به صورت

۱-۲
$$Cc(vi) = \frac{Ntriangles(vi)}{Ntriples(vi)}$$

می‌باشد. که $Ntriangles(vi)$ تعداد مثلث‌هایی می‌باشد که نود v_i را به عنوان یک نود دارند و$Ntriples(vi)$ تعداد سه‌تایی‌هایی می‌باشد که نود v_i یکی از نودهای آن باشد. ضریب خوشه‌بندی تقسیم تعداد لینک‌های بین نودهای همسایه‌ی یک نود بر تعداد لینک‌هایی است که پتانسیل به وجود آمدن بین نودهای همسایه‌های v_i را دارند. ضریب خوشه‌بندی یک گراف میانگین ضریب خوشه‌بندی تمام نودهای شبکه می‌باشد.

۱-۳
$$Cc(G) = \frac{1}{|V|}\sum_{vi} Cc(vi)$$

شبکه‌های پیچیده معمولا تمایل به داشتن ضریب خوشه‌ی بالا دارند.

۴- میانگین فاصله: فاصله‌ی بین دو نود طول کوتاه‌ترین مسیر بین دو نود در گراف می‌باشد. میانگین فاصله، میانگین تمام کوتاه‌ترین مسیرها در گراف می‌باشد. در شبکه‌های پیچیده این مقدار معمولا کوچک می‌باشد. برای گراف بدون وزن G با N نود میانگین فاصله به صورت زیر محاسبه می‌شود:

۱-۴
$$Distance_{avg}(G) = \frac{2}{N.(N-1)}\sum_{vi,vj} dist(vi, vj)$$

۵- قطر: قطر یک گراف طول بزرگترین مسیر کوتاه بین هر دو نودی می‌باشد.

$$Diameter(G) = \max(\{dist(vi, vj) \forall\ vi, vj\ \in V\}) \quad ۱-۵$$

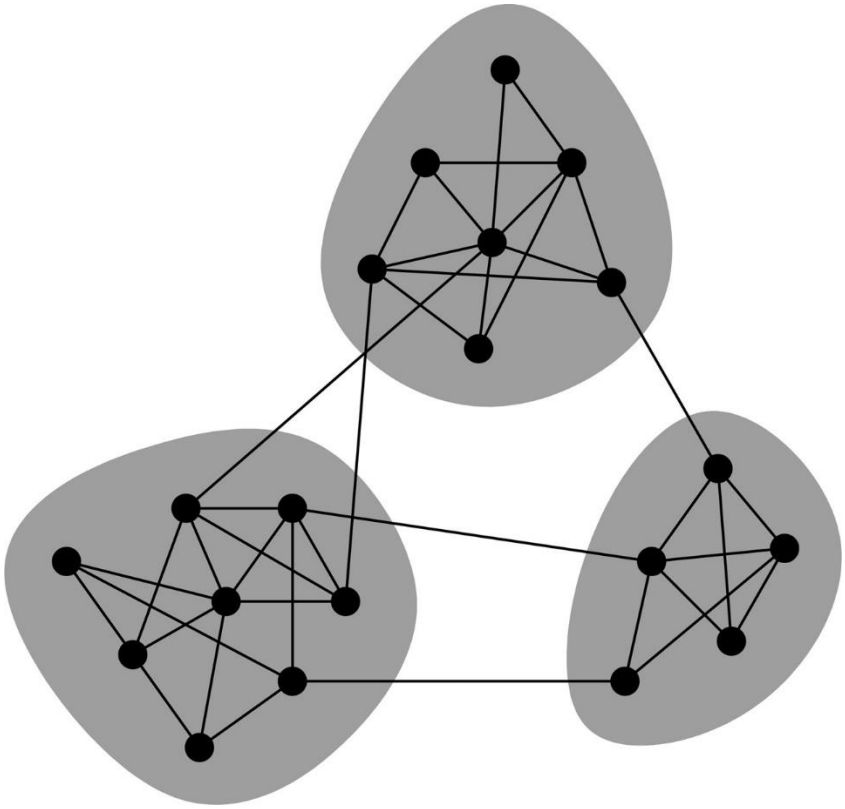
در شبکه‌های پیچیده قطرها معمولا کوتاه می‌باشند. اگر شبکه پیوسته باشد محاسبه قطر طبق تعریف انجام می‌شود. اگر گراف از چند زیرگراف پیوسته تشکیل شده باشد، قطر گراف میانگین قطر زیرگراف‌ها می‌باشد. بیشتر شبکه‌های پیچیده معمولا میانگین ضریب خوشه‌بندی بالا و میانگین فاصله و قطر کوتاهی دارند.

۶- تراکم: تراکم یک شبکه نسبت لینک‌های موجود به تمام لینک‌های ممکن در یک گراف می‌باشد.

$$Density(G) = \frac{2|E|}{|V| * (|V| - 1)} \quad ۱-۶$$

بیشتر شبکه‌های پیچیده تراکم اندکی دارند. به معنی دیگر معمولا تنک می‌باشند.

۷- ساختار انجمنی: شبکه‌های پیچیده تمایل به داشتن خوشه‌هایی از نودها در قالب انجمن‌ها دارند. انجمن‌ها زیرگراف‌هایی در شبکه می‌باشند که نودها در آن‌ها ارتباط‌های شبیه به هم دارند. نودهای یک انجمن بیشتر بین یکدیگر لینک دارند تا با نودهای یک انجمن دیگر.



شکل ۳) ساختار انجمنی در شبکه‌های پیچیده http://www.pnas.org/content/103/23/8577

انجمن‌ها می‌توانند با یکدیگر اشتراکاتی داشته باشند. در [5] معیاری به نام modularity برای ارزیابی انجمن‌ها ارائه شده است.

در [6] توضیحات بیشتری از ویژگی‌های مختلف آماری انواع مختلف شبکه‌ها معرفی شده است.

۳. مدل‌سازی شبکه:

مدل‌سازی شبکه معمولا به منظور نشان دادن شبکه در قالب فرمول‌ها می‌باشد تا بتوان با اعمال ریاضی ویژگی‌های آن را تحلیل و بررسی کرد. همچنین از این ابزار می‌توان برای پیش‌بینی برخی از ویژگی‌های شبکه نیز استفاده کرد. در زیر سه مدل شناخته شده معرفی می‌شوند:

۱- گراف‌های تصادفی[۱]: گراف‌های تصادفی گراف‌هایی هستند که لینک بین نودهای ترتیب خاصی در آن ندارد. بدین معنا که می‌توان این گراف را با قرار دادن لینک‌های تصادفی بین نودها ساخت. اولین مدل احتمالاتی برای تولید گراف تصادفی در [7, 8] ارائه شد. در این روش دو مدل ارائه شد. در مدل اول شبکه‌ای با n نود و m یال ساختند. این روش از n نود جدا از هم شروع می‌کند. به صورت تصادفی دو نود را انتخاب می‌کند و بین آن‌ها یالی قرار می‌دهد تا زمانی که تعداد یال‌های m شود. در روش دوم برای ساخت گراف بین هر دو نود با احتمال $0 < p < 1$ یالی ایجاد می‌کند. با تکرار این روش گراف‌های مختلفی ساخته می‌شود که هر گراف تعداد یال‌های متفاوتی دارند. یک گراف با m یال با احتمال $p^m (1-p)^{C-m}$ ایجاد

[۱] random graph

می‌شود که $C = \frac{n(n-1)}{2}$ تعداد کل لینک‌های ممکن در گراف می‌باشد. برخی از ویژگی‌های گراف‌های تصادفی در [9] بررسی شده است. که از جلمه‌ی آن‌ها می‌توان به موارد زیر اشاره کرد:

۱- اگر $p > 1/n$ باشد میانگین درجه $k\ avg = 1$ و اگر $p > \ln(N)/n$ باشد اکثر گراف‌های تصادفی پیوسته می‌شوند.

۲- اگر n خیلی بزرگ باشد $k\ avg \simeq p.n$ و درجه توزیع $P\ (k)$ با توزیع پواسن به صورت زیر تخمین زده می‌شود.

۱-۷
$$P(k) = k_{avg}^{k}.\frac{e^{-k_{avg}}}{k!}$$

به همین خاطر به این گراف‌ها، گراف تصادفی پواسنی نیز گفته می‌شود.

۳- قطر یک گراف تصادفی بین مقادیر اندک $\frac{\ln(n)}{\ln(p.n)} \approx \frac{\ln(n)}{\ln(k_{avg})}, p.n \to inf$ تغییر می‌کند. [9]

۴- ضریب خوشه‌بندی در گراف‌های تصادفی برابر p یا k_{avg}/n می‌باشد [10, 11]. دلیل این امر این است که طبق تعریف وجود لینک بین نودها مستقل از یکدیگر می‌باشد. بنابراین احتمال این که بین دو نود یال جدید برقرار شود به شرط این که همسایه‌ی مشترک داشته باشند بالاتر نمی‌رود.

۲- گراف‌های دنیای کوچک[۱]: این گراف‌ها در [10] معرفی می‌شوند. در این تحقیق شبکه‌های بسیاری ارزیابی می‌شوند و مشاهده می‌شود در بسیاری از شبکه‌های واقعی علاوه بر اینکه میانگین فاصله‌ی کوتاهی دارند، ضریب خوشه‌بندی بالاتری از مقدار مورد انتظار در حالت تصادفی دارند. گراف دنیای کوچک گرافی است که بسیاری از نودها ممکن است همسایه نباشند، ولی بسیاری از نودها با تعداد گام‌های اندکی می‌توانند به بسیاری از دیگر نودها برسند. این گراف‌ها میانگین کوتاه‌ترین فاصله‌ی کوچکی دارند و فاصله‌ی d بین هر دو نود تصادفی به نسبت $\log(n)$ که n تعداد نودها می‌باشد زیاد می‌شود. به عبارتی $d \propto \ln(n)$. ساخت گراف با n نود و l یال برای هر نود آغاز می‌شود. هر یال با احتمال p می‌تواند به یک نود تصادفی دیگر وصل شود. اگر $p=1$ باشد این گراف یک گراف تصادفی می‌شود. با وجود اینکه گراف‌های دنیای کوچک نسبت به گراف‌های تصادفی به دنیای واقعی نزدیک‌تر می‌باشند ولی محدودیت‌هایی دارد. اول این‌که گراف دنیای کوچک پویایی شبکه‌های واقعی را ندارد. ثانیا توزیع درجه‌ی نودهای یک گراف معمولا زنگوله‌ای نیست. بلکه معمولا از قانون قدرت پیروی می‌کند که وجود نودهای هاب در شبکه را نمایش می‌دهد. هرچند برخی ویژگی‌های گراف دنیای کوچک در گراف‌هایی همچون شبکه غذایی، شبکه‌ی اینترنت [12]، شبکه توزیع نیرو [10]، شبکه‌ی حمل و نقل، شبکه‌های بیولوژی [13] و شبکه‌ی همکاری‌های علمی مشاهده می‌شود.

۳- گراف‌های بدون مقیاس[۲]: مدل‌هایی که توزیع درجه در آن‌ها از توزیع پواسنی فاصله دارد، مدل‌های بدون مقیاس نامیده می‌شوند. در بسیاری از شبکه‌های واقعی توزیع درجه از مدل زنگوله‌ای پیروی نمی‌کند. بلکه از مدل قانون قدرت پیروی می‌کند. $P\ (k) \sim c\Delta k^{-\gamma}$ که k درجه‌ی نودها، c عددی ثابت و و γ مقداری مثبت است که بین دو سه تغییر مقدار

---

۱ Small world graphs

۲ scale-free

می‌کند. در قانون قدرت تعداد بسیار زیادی از نودها درجه‌ی پایینی دارند و تعداد اندکی از نودها که هاب نامیده می‌شوند درجه‌ی بسیار بالایی دارند. هاب‌ها نقش بسیار به سزایی در تداوم، گسترش و پیوستگی شبکه دارند.

۴- تحلیل شبکه:

در [14] تحلیل شبکه‌ها به دسته‌ی کلی تقسیم شده است:

۱- تحلیل ساختاری: این دسته از تحلیل‌ها تنها از اطلاعات ساختاری شبکه استفاده می‌کنند. دانش بیشتری از ویژگی‌های نودها و لینک‌ها وجود ندارد. این تحلیل‌ها شامل تحلیل‌های آماری شبکه، تشخیص انجمن، طبقه‌بندی نودها و یا برچسب‌گذاری بر روی نودها، پیش‌بینی لینک و نمایش‌سازی شبکه می‌باشد. در [14] بررسی اجمالی‌ای از رفتار شبکه‌ها انجام شده است. برخی از تحلیل‌های مبتنی بر ساختار شبکه در تحقیقات زیر قابل مشاهده می‌باشد. [22 -15]

۲- تحلیل مبتنی بر محتوا: این روش‌ها از ویژگی‌ها و محتوای شبکه برای تحلیل نیز بهره می‌برند. ویژگی‌هایی که برای استخراج و تحلیل آن‌ها از روش‌های داده‌کاوی، متن‌کاوی و ... استفاده می‌شود. در تحقیقات [25 -23]

نشان داده شده است که استفاده از محتوای شبکه‌ها تا چه اندازه‌ای می‌تواند دانش با ارزش برای تحلیل شبکه فراهم کند.

تقسیم‌بندی‌های مختلفی از سطوح عملیات تحلیل شبکه‌ها ارائه شده است. در [26] این تقسیم‌بندی ارائه شده است:

عملیات در سطح نود: این سطح شامل عملیات پیدا کردن نودهای مهم با توجه به نودهای دیگر شبکه می‌باشد. مساله‌ی مرکزیت (centrality) در این سطح قرار می‌گیرد.

عملیت در سطح دوتایی: عملیات در این سطح دو نود را شامل می‌شود. برخی از عملیات شامل پیدا کردن فاصله بین دو نود و یا احتمال ایجاد لینک بین دو نود (پیش‌بینی لینک) می‌شود.

عملیات در سطح سه‌تایی: در این سطح دسته‌های سه‌تایی نودها بررسی می‌شوند. از جمله عملیات در این سطح می‌توان به پیدا کردن ضریب خوشه‌ی محلی اشاره کرد.

عملیات در سطح زیرمجموعه‌ها: این سطح گروهی از نودها را بررسی می‌کند. پیدا کردن انجمن‌ها در این سطح قرار می‌گیرد.

عملیات در سطح شبکه: این عملیات کل شبکه را برای استخراج برخی ویژگی‌ها تحلیل می‌کنند. مانند پیدا کردن قطر، تراکم و پیوستگی گراف.

از جمله برخی عملیات‌های مهم در تحلیل گراف می‌توان به موارد زیر اشاره کرد:

۱- مرکزیت: یکی از عملیات‌های اولیه در تحلیل شبکه‌ها پیدا کردن نودهای مهم در شبکه می‌باشد. این نودها نقش مهمی در توزیع اطلاعات در شبکه و تاثیر بر دیگر نودها دارند. مرکزیت یک نود اهمیت نسبی یک نود در شبکه می‌باشد [27] گونه‌های مختلفی از مرکزیت وجود دارد، از جمله مرکزیت درجه، مرکزیت نزدیکی[۱] ، مرکزیت بینی[۲] ، مرکزیت مقادیر ویژه[۳] .

مرکزیت درجه ساده‌ترین معیار مرکزیت می‌باشد که استفاده از آن در تحلیل شبکه بسیار متداول می‌باشد. مرکزیت درجه برای یک نود اندازه می‌گیرد که یک نود به چه تعداد نود دیگر متصل است . در گراف جهت‌دار درجه ورودی و درجه خروجی برای نمایش تعداد یال‌های ورودی و خروجی استفاده می‌شوند. معمولا این مقدار با بیشتر تعداد لینک ممکن برای یک نود در گراف نرمالیزه می‌شود. برای شبکه‌ای که N نود دارد درجه مرکزیت نود v_i به صورت:

۱-۷
$$C_D(v_i) = \frac{\deg(v_i)}{N-1}$$

پیچیدگی محاسباتی مرکزیت درجه $O(N)$ می‌باشد که باعث شده استفاده از آن در گراف‌های بزرگ مناسب باشد.

نوع دیگر مرکزیت، مرکزیت نزدیکی[۴] [27] می‌باشد که فاصله‌ی یک نود از تمام نودهای دیگر در شبکه اندازه می‌گیرد. یک نود مرکزیت دارد اگر به تمام نودهای دیگر نزدیک باشد. این معیار از طریق معکوس مجموع فاصله‌های یک نود از دیگر نودها حساب می‌شود.

۱-۸
$$C_c(v_i) = \left[\sum_{j=1}^{N} dist(v_i, v_j)\right]^{-1}$$

بیشترین مقدار ممکن برای مرکزیت نزدیکی برای یک نود $(N-1)^{-1}$ می‌باشد زمانی رخ می دهد که تمام نودهای دیگر به این نود متصل باشند. کمترین مقدار آن می‌تواند صفر باشد. زمانی که نود به نود دیگری وصل نیست. در مدل استاندار مقدار مرکزیت نزدیکی بین ۰ و ۱ می‌باشد.

۱-۹
$$C_c(v_i) = \frac{N-1}{\sum_{j=1}^{N} dist(v_i, v_j)}$$

پیچیدگی محاسباتی این روش $O(N\log(N) + M)$ می‌باشد.

---

[۱] closeness
[۲] betweenness
[۳] eigenvector
[۴] closeness centrality

معیار دیگر مرکزیت بینی[1] [27, 28] می‌باشد. در مرکزیت بینی تعداد دفعاتی را که یک نود در کوتاه‌ترین مسیر بین هر دو نود قرار می‌گیرد، محاسبه می‌شود. در شبکه‌های اجتماعی ارتباط میان دو نود می‌تواند وابسته به دیگر نودها باشد. به خصوص نودهایی که در مسیر بین دو نود قرار می‌گیرند. بنابراین یک نود مرکزیت دارد اگر در تعداد بیشتری از کوتاه‌ترین مسیرها در شبکه ظاهر شود.

۱-۱۰

$$C_B(v_i) = \sum_{i \neq j \neq k} \frac{|spaths(v_i, v_k|v_i)|}{|spaths(v_j, v_k)|}$$

بیشترین مقدار مرکزیت بینی $\frac{(N-1)(N-2)}{2}$ می‌باشد. بنابراین مدل استاندارد آن به صورت زیر می‌شود:

۱-۱۱

$$C_B(v_i) = \left(\frac{2}{(N-1)(N-2)}\right) \sum_{i \neq j \neq k} \frac{|spaths(v_i, v_k|v_i)|}{|spaths(v_j, v_k)|}$$

که $spaths(v_j, v_k|v_i)$ مجموعه تعداد کوتاه‌ترین مسیرهای بین دو نود v_j و v_k می‌باشد که شامل v_i می‌باشند. پیچیدگی محاسباتی این روش $O(N.M + N^2 \log(N))$ می‌باشد. استفاده از این روش برای گراف‌های بزرگ هزینه‌بر می‌باشد.

نوع دیگر مرکزیت، مرکزیت مقادیر ویژه می‌باشد که تاثیر یک نود در شبکه را اندازه‌گیری می‌کند. این روش مقداری نسبی به تمامی نودهای شبکه نسبت می‌دهد. بر این اساس که لینک‌های متصل به نودهای با امتیاز بالا نقش بیشتری در امتیاز نودها دارند تا لینک‌هایی که به نودهای با امتیاز پایین متصل می‌باشند.

۱-۱۲

$$C_E = \left(\frac{1}{\lambda}\right) \sum_{u \in \Gamma} C_E(u)$$

پیچیدگی محاسباتی این روش $O(N^2)$ می‌باشد.

یکی دیگر از مفاهیم مشابه مرکزیت برای مشخص کردن اهمیت نودها در شبکه، روش اعتبار[2] می‌باشد. این روش در گراف‌های جهت‌دار استفاده می‌شود. یک نود معتبر است اگر مقصد بسیاری از لینک‌ها باشد. به این معنی که درجه‌ی ورودی بالایی دارد. [27]. روش PageRank گوگل مثالی از این روش می‌باشد. این روش تحلیل لینک وزنی به هایپرلینک‌های یک صفحه‌ی وب بر اساس اهمیت صفحه‌های دیگر که به این صفحه لینک داده‌اند نسبت می‌دهد. به نوعی مرکزیت مقادیر ویژه را در نظر می‌گیرد.

---

[1] betweenness centrality

[2] Prestige

۲- تشخیص انجمن[۱]: انجمن‌ها مجموعه‌ای از نودها می‌باشند که ویژگی‌های مشترکی دارند و یا نقش مشابهی در شبکه ایفا می‌کنند[29]. ساختار انجمنی در بسیاری از شبکه‌های واقعی مانند ارتباط پروتئین‌ها [30, 31]، انجمن‌ها در شبکه‌های اجتماعی [32]، شبکه‌ی وب [33] و ... دیده می‌شود. تشخیص انجمن تلاش می‌کند ناحیه‌های چگال در شبکه را پیدا کند که نودها در آن ویژگی‌های مشترک و رفتارهای مشابه دارند. این مفهوم بسیار به مفهوم خوشه‌بندی نزدیک است. استفاده‌ی بسیاری در سیستم‌های توصیه‌گر، تحلیل شبکه وب، دسته‌بندی نودها و... دارد. هدف اصلی تشخیص انجمن پیدا کردن مادول‌ها و ساختار سلسله مراتبی آن‌ها با استفاده از تحلیل ساختار گراف می‌باشد. همچنین می‌توان دانش حاصل از محتوای عناصر شبکه را در تشخیص انجمن به کار برد و کیفیت خروجی را بالا برد.

۳- پیش‌بینی لینک: بیشتر تحقیقات بر روی شبکه‌های پیچیده برای پیدا کردن الگوی لینک‌های بین نودها می‌باشد. بنابراین پیش‌بینی لینک یکی از موضوعات مهم در تحلیل شبکه‌ها می‌باشد. لینک‌ها ارتباط‌های مختلفی بین دو نود را نشان می‌دهند. این ارتباط می‌تواند روابط دوستی در شبکه‌های اجتماعی، روابط همکاری در شبکه‌ی همکاری‌های علمی و ... باشد. روش پیش‌بینی لینک تلاش دارد با مشاهده‌ی شبکه در زمان t، لینک‌های نادیده شده در زمان t یا لینک‌هایی که در زمان t+k ایجاد می‌شوند را پیش‌بینی کند. پیدا کردن نودهای از قلم افتاده موجب نزدیک شدن به ساختار واقعی شبکه و بهبود تحلیل اطلاعات می‌شود. پیش‌بینی لینک می‌تواند نقش مهمی در عملیات‌های تحلیلی دیگر همچون تشخیص انجمن‌ها [20,35] و یا تشخیص نودهای مهم و تاثیرگذار [36] داشته باشد. در تحقیق [35] برای پیدا کردن انجمن‌ها از معیارهای شباهت نودها استفاده شده که بیشتر در مساله‌ی پیش‌بینی لینک به آن‌ها پرداخته می‌شود. در این مقاله نشان داده شده است که استفاده از معیارهای پیش‌بینی لینک در روش‌های تشخیص انجمن‌ها بهبودی قابل ملاحظه‌ای را نتیجه داده است.

۴- هاب‌ها و کلیک‌ها[۲]: کلیک در شبکه گروهی از نودها می‌باشند که زیرگراف کاملی می‌سازند. مساله‌ی یافتن اینکه آیا در گراف کلیک وجود دارد NP-hard میباشد [26]. بیشینه کلیک بزرگترین کلیک ممکن در یک شبکه می‌باشد. از کلیک‌ها در عملیات‌های دیگری همچون تشخیص انجمن و یا پیش‌بینی لینک نیز استفاده می‌شود [20]. در این تحقیق نویسنده نشان داده است که لینک‌ها تمایل به ساخت کلیک در شبکه دارند و از آن در پیش‌بینی لینک استفاده کرده‌اند. در این تحقیق سه رفتار مشاهده شده است. ۱- لینک‌ها تمایل به ساخت کلیک در شبکه دارند. ۲- لینک‌ها بیشتر تمایل دارند کلیک‌های بزرگ بسازند تا کلیک‌های کوچک. ۳- یک لینک تمایل دارد در چندین کلیک نقش داشته باشد. با استفاده از این رفتارها نویسنده ایجاد لینک در انجمن‌ها را بررسی کرده است.

۵- شبکه‌ی مقالات علمی:

شبکه‌ی مقالات علمی اطلاعات بسیاری مرتبط با انتشارات علمی در زمینه‌های مختلف تحقیقاتی دارند. این شبکه‌ها ساخته شده از محققان، مقالات، ژورنال‌ها و ... دارند. همچنین ممکن است اطلاعاتی مربوط به متن مقاله در قالب چکیده و کلمات کلیدی داشته باشند. یکی از زمینه‌های تحقیقاتی مهم در تحلیل شبکه‌ی مقالات علمی، تحلیل همکاری‌های علمی می باشد. از مقالات علمی شبکه‌های مختلفی می‌توان ساخت. شبکه‌ی همکاری‌های علمی که در آن نویسندگان نودهای شبکه می‌باشند و بین دو نویسنده در شبکه لینک وجود دارد اگر حداقل یک مقاله‌ی مشترک داشته باشند. شبکه‌ی دیگر شبکه‌ای می‌باشد که

---

۱ Community

۲ clique

در آن نودها مقالات می باشند و بین دو نود لینک وجود دارد اگر نویسنده‌ی مشترک داشته باشند و یا اینکه به یکدیگر اشاره (cite) کرده باشند. همچنین می‌توان از مقالات گراف دو بخشی نویسنده مقاله ساخت. در این شبکه دو دسته نود وجود دارد. نودهای مقالات و نودهای نویسندگان. بین یک نود نویسنده و یک نود مقاله لینک وجود دارد اگر آن نویسنده در نوشتن مقاله نقش داشته باشد. همچنین لینک می‌تواند وجود داشته باشد اگر نویسنده به مقاله اشاره کرده باشد.

مراجع

[1] Ernesto Estrada. The Structure of Complex Networks: Theory and Applications. Oxford University Press, 2011.

[2] M. E. J. Newman. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Science of the United States (PNAS), 101:5200–5205, 2004a.

[3] A. Barrat, M. Barthélemy, and A. Vespigani. Modeling the evolution of weighted networks. Physical Review E 70:066149, 2004.

[4] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. Nature, 393(6684):440–442, 1998.

[5] M.E. J. Newman and M M. Girvan. Finding and evaluating community structure in networks. Physics review E, 69:026113:1–022613:15, 2004.

[6] Mary McGlohon, Leman Akoglu, and Christos Faloutsos. Statistical properties of social networks. In Social Network Data Analytics, pages 17–42. Springer, 2011.

[7] Béla Bollobás. Random graphs. Cambridge University Press, 2 edition, 2001. ISBN 0521797225.

[8] Paul Erdős and Alfréd Rényi. On random graphs. Publicationes Mathematicae Debrecen,6:290–297, 1959.

[9] Fan Chung and Linyuan Lu. The diameter of random sparse graphs. Advances in Applied Math, 26(4):257–279, 2001.

[10] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. Nature, 393(6684):440–442, 1998.

[11] Mark Newman. Random graphs as models of networks. Handbook of Graphs and Networks, pages 35–68, 2005. doi: 10.1002/3527602755.ch2.

[12] Lada Adamic, Orkut Buyukkokten, and Eytan Adar. A social network caught in the Web. First Monday, 8(6), 2003.

[13] Albert-László Barabási and Z. N. Oltvai. Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 5:101–113, 2004.

[14] Charu C. Aggarwal. Social Network Data Analytics, chapter An introduction to social network data analytics. Springer, 2011.

[15] Yong-Yeol Ahn, Seungyeop Han, Haewoon Kwak, Sue Moon, and Hawoong Jeong. Analysis of topological characteristics of huge online social networking services. In Proceedings of the 16th international conference on World Wide Web, pages 835–844, 2007.

[16] Nesserine Benchettara, Rushed Kanawati, and Céline Rouveirol. Supervised machine learning applied to link prediction in bipartite social networks. In International Conference on Advances in Social Network Analysis and Mining, ASONAM 2010, pages 326–330, 2010a.

[17] Zan Huang. Link prediction based on graph topology: The predictive value of the generalized clustering coeffecient. In Proceedings of LinkKDD'06, Philadelphia, Pennsylvania, 2006.

[18] Lin Li, Bao-Yan Gu, and Li Chen. The topological characteristics and community structure in consumer-service bipartite graph. In Jie Zhou, editor, Complex (1), volume 4 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pages 640–650. Springer, 2009. ISBN 978-3-642-02465-8.

[19] David Liben-Nowell and Jon M Kleinberg. The link-prediction problem for social networks. JASIST, 58(7):1019–1031, 2007.

[20] Zhen Liu, Jia-Lin He, and Jaideep Srivastava. Cliques in complex networks reveal link formation and community evolution. CoRR, 2013.

[21] M. E. J. Newman. Communities, modules and large-scale structure in networks. Nature Physics, 8:25–31, 2012.

[22] Zied Yakoubi and Rushed Kanawati. Licod: A leader-driven algorithm for community detection in complex networks. Vietnam Journal of Computer Science, 1 (4):241–256, 2014.

[23] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki. Link prediction using supervised learning. In Workshop on link analysis, Counter-terrorism and security, SIAM Data Mining Conference, Bethesda, MD, 2006.

[24] Alexandrin Popescul and Lyle H Ungar. Cluster-based concept invention for statistical relational learning. In Won Kim, Ron Kohavi, Johannes Gehrke, and William DuMouchel, editors, KDD, pages 665–670. ACM, 2004. ISBN 1-58113-888-1.

[25] Chao Wang, Venu Satuluri, and Srinivasan Parthasarathy. Local probabilistic models for link prediction. In Yong Shi and Christopher W Clifton, editors, Seventh IEEE International Conference on Data Mining (ICDM), pages 322–331. IEEE, October 2007.

[26] Stanley Wasserman and Katherine Faust. Social Network Analysis: Methods and Applications, chapter Social Network Analysis in the Social and Behavioral Sciences, pages 3–27. Number 8. Cambridge University Press, 1994.

[27] Katherine Faust and Stanley Wasserman. Centrality and prestige: A review and synthesis. Journal of Quantitative Anthropology, 4(1):23–78, 1992.

[28] L.C. Freeman. A set of measures of centrality based upon betweenness. In Sociometry 40, pages 35–41, 1977.

[29] S. Fortunato. Community detection in graphs. Physics Reports, 486(3-5):75–174, 2010.

[30] Roger Guimera and Luis A. Nunes Amaral. Functional cartography of complex metabolic networks. Nature, 433(7028):895–900, 02 2005. URL http://dx.doi.org/10.1038/nature03288.

[31] G. Palla, I. Derônyi, I. Farkas, and T. Vicsek. Uncovering the overlapping modular structure of protein interaction networks. FEBS Journal, 272:434, 2005.

[32] Linton C. Freeman. The development of social network analysis: A study in the sociology of science, volume 1. Empirical Press Vancouver, 2004.

[33] Yon Dourisboure, Filippo Geraci, and Marco Pellegrini. Extraction and classification of dense communities in the web. In Proceedings of the 16th international conference on World Wide Web, pages 461–470. ACM, 2007.

[34] Roger Guimerà, Stefano Mossa, Adrian Turtschi, and Luis A. Nunes Amaral. The world-wide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proceedings of the National Academy of Sciences, 102(22):7794–7799, 2005.

[35] Bowen Yan and Steve Gregory. Detecting communities in networks by merging cliques. In Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International

Conference on, volume 1, pages 832–836. IEEE, 2009.

[36] K. Subbian and P. Melville. Supervised rank aggregation for predicting influence in networks. In Proceedings of the IEEE Conference on Social Computing (SocialCom-2011)., Boston, October 2011.