به نام پاک آفریدگار

پروژه دوم درس شبیه‌سازی سیستم‌های کامپیوتری

دانشکده مهندسی و علوم کامپیوتر، دانشگاه شهید بهشتی، نیمسال دوم ۴۰۰-۴۰۱

بررسی تاب آوری و استحکام شبکه‌ها در برابر خرابی‌های ایستا و پویای گره‌ها

مدرس: فرشاد صفایی

۱- مقدمه

میدانیم در شبکه‌ها، افراد و سازمان‌های مختلف به‌واسطه علایق، سلایق، عقاید و منافع مشترک با یکدیگر در ارتباط هستند و می‌توان آن‌ها را به مانند گراف‌هایی تصور کرد که گره‌های موجود در آن که توسط یک یا چند نوع خاص از وابستگی (پیوند یا یال ارتباطی) به یکدیگر اتصال یافته‌اند. شبکه‌ها بخش مهمی از دنیای پیرامون ما را شکل میدهند و در عملکرد روزانه ما دخالتی تام و تمام دارند؛ با این همه، ممکن است در برابر گسترهٔ وسیعی از چالش‌ها و حملات از جمله حملات هوشمند، اشکالات نرم‌افزاری، سخت‌افزاری، و اشتباهات انسانی، تاب‌آوری (robustness) را برای سرویس‌دهی مورد انتظار نمایش ندهند. بدین سیاق، انگیزش اصلی بسیاری از پژوهشها به طراحی شبکه‌هایی اختصاص یافته که تاب آور باشند یعنی کاربرد حیاتی شبکه‌هایی مانند اینترنت، نیروگاه‌های برق، سیستم‌های حمل و نقل، شبکه‌های سنسور، شبکه‌های لجستیکی و امثالهم در برابر خرابیها بررسی و تحلیل گردد. بدین ترتیب، استحکام در شبکه‌ها و زیرساخت‌های آن، یکی از ویژگی‌های حیاتی و مهم به شمار میرود و در طی سالیان اخیر به یکی از زمینه‌های پژوهشی جذاب و رو به رشد تبدیل شده است. این زمینهٔ پژوهشی به دنبال آن است تا راه‌کارها و مکانیزم‌هایی را جهت بهبود اتصال‌پذیری (همبندی) شبکه‌ها در برابر خرابی‌ها جستجو کند.

۲- شرح پروژه

در گذر سال‌ها، مسایل بهینه‌سازیِ گراف در ادبیات پژوهش مطرح گشته‌اند که در طراحی و تحلیل شبکه‌ها نیز به چشم می‌خورند. یک فرض همیشگی در بیشتر این پژوهشها و مفروضاتِ مدل‌سازی تحلیلی این است که خرابی‌ها به گونه‌ای هستند که شبکه را ناهمبند نمیسازند. اکنون هدف ما در انجام این پروژه این است که احتمال ناهمبندی را در دو حالت ایستا و پویای شبکه که دارای تعدادی گرهٔ پردازشی است مورد بررسی و تحلیل قرار داده و نشان دهیم که چگونه الگوهای ایستا و پویای خرابی‌هایی از نوع گره ممکن است بر استحکام و تاب‌آوری شبکه‌ها تأثیر بگذارند.

تعریف ۱: منظور از ایزوله شدن شبکه یعنی اینکه گره یا گره‌هایی داشته باشیم که دارای درجه صفر باشند. احتمال پیشامد چنین گره‌هایی را با φ یعنی پیشامد ایزوله شدن شبکه نمایش میدهیم؛ همچنین، متغیر تصادفی T بیانگر مدت زمان ایزوله شدن است؛ یعنی ماکزیمم زمانی که گره در سیستم وجود دارد تا سرانجام مجبور به ترک شبکه گردد. به بیان ریاضی میتوان نوشت

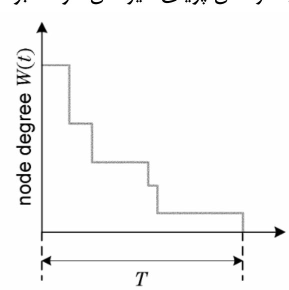
$$T = \max\{t > 0 : W(t) = 0 \mid W(0) = k\} \qquad (۱)$$

که نشاندهندهٔ یک احتمال شرطی است و W(0)=k درجه اولیه گره را نشان میدهد که برابر با k است. بدین ترتیب، E[T] متوسط زمان تا ایزوله شدن و φ احتمال وقوع چنین پیشامدی را در طی طول عمر کاربران نشان میدهد که یکی از خواسته‌های ما در انجام این شبیه‌سازی است.

تعریف ۲: منظور از الگوی خرابی ایستای گره این است که گره‌ها با یک احتمال مشخص و مفروضی مانند p دچار خرابی میگردند که این احتمال برای هر گره از سایرین مستقل است

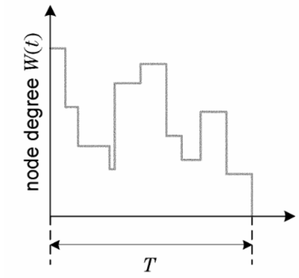
تعریف ۳: منظور از الگوی خرابی پویای گره این است که هر گره دارای یک طول عمر است که با یک متغیر تصادفی مشخص میگردد که از یک توزیع احتمال مشخص قابل حصول است

تعریف ۴: مدل خرابی پویا خود دارای دو زیرشاخه است؛ یکی مدل غیرفعال (passive) و دیگری مدل فعال (active). در مدل غیرفعال، کاربران موجود در شبکه (یعنی گره‌ها) به طور پیوسته سیمبندی (rewire) نمیشوند؛ یعنی در این مدل، یک گره در صورت از دست دادن همسایه‌هایش (به دلیل انقضای طول عمر آن‌ها) تلاشی در یافتن همسایه مناسب و جایگزینی آن‌ها ندارد. چنانچه W(t) درجه لحظه‌ای گره در زمان t باشد، نمودار زیر بیانگر مدل پویای غیرفعال خواهد بود. این منحنی در حقیقت تکامل درجه گره را در طی زمان نشان میدهد.



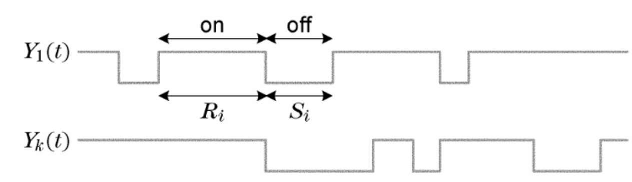
(a) passive model

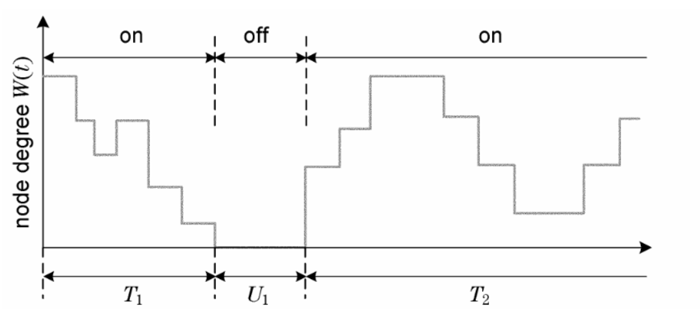
در مقابل، در مدل پویای فعال، گره آسیب دیده (affected) – یعنی گره‌ای که همسایگانش را از دست داده– قادر به ترمیم خویش است و میتواند با یک جستجو، همسایهٔ مناسبی را پیدا و جایگزین همسایه های از دست رفته کند. اگر S بیانگر متغیر تصادفی زمان جستجوی برای یافتن همسایه‌های مناسب باشد، E[s] متوسط این زمان جستجو را نشان میدهد که میتواند از توزیع یکنواخت، نمایی، نرمال بریده شده و ... پیروی کند. نمودار زیر این مدل را به شکل مفهومی نشان داده است.



(b) active model

یک نکته بسیار جالب این است که در مدل پویای فعال، اگر یک گره دارای درجه i (یعنی i همسایه) باشد، این همسایگان در زمان t یا زنده هستند (در وضعیت on) یا اینکه به دلیل انقضای طول عمر خراب شده‌اند (در وضعیت off). بدین ترتیب، ما با یک پراسیچر خرابی/جایگزینی در هر همسایه رو به رو هستیم که میتواند با یک فرآیند on/off برای مثال $Y_i(t)$ بیان گردد؛ که اگر مقدار ۱ داشته باشد به معنای زنده بودن همسایه این گره در لحظه t و اگر برابر صفر باشد یعنی این همسایه معیوب شده است. لذا، وقتی گره‌ای، k همسایه داشته باشد به معنای داشتن یک فرآیند $Y_1(t), Y_2(t), \dots, Y_k(t)$ است. بدیهی است درجه یک گره در لحظه t برابر با مجموع این k فرآیند است یعنی $W(t)=\sum_{i=1}^{k} Y_i(t)$ که عملاً به معنای اصل برهمنهی فرآیندهای برنولی (on/off) مدوله شده با یک زنجیره مارکوف یعنی همان MMP است که در کلاس درس به آن اشاره داشتیم و برای تولید فرآیندهای دُم-کلفت یا خودمتشابه (رگباری) کاربرد دارد. پس به طور خلاصه؛ در مدل پویای فعال، دنباله یا توزیع درجه از یک توزیع حافظه‌دار یا دُم-کلفت (قانون قدرت) پیروی میکند و میتواند بیانگر تاب آوری و استحکام بیشتر شبکه در مقایسه با مدل پویای غیرفعال باشد که عملاً مُهر تاییدی بر نتایج شبیه‌سازی و شهودی با مدل تحلیلی است. شکل زیر توضیحات بالا را برای درک بیشتر به صورت بصری به تصویر کشیده است.





۳- مراحل انجام شبیه‌سازی

-------------------------------- فاز اول --------------------------------

۳-۱. بررسی تاب‌آوری شبکه تحت مدل خرابی ایستای گره‌ها

همانطور که در مقدمه و در بالا اشاره کردیم، منظور از مدل ایستای طول عمر آن است که گره‌ها مستقل از یکدیگر با یک احتمال مشخص p میتوانند دچار خرابی گردند. احتمال ایزوله شدن گره‌ای مانند i پس از چندین بار تلاش در صورتی که درجه‌اش ki باشد (یعنی تعداد همسایگانی که بایستی معیوب شوند تا گره i ایزوله گردد) دارای یک توزیع هندسی است. چنانچه متغیر تصادفی Z بیانگر تعداد گره‌های ایزوله باشد، میتوان احتمال همبندی گراف G (عدم وجود گره‌های ایزوله) را به صورت زیر تعریف کرد

$$P\{G \text{ is connected}\}=\text{P}\{\text{Z}=0\} \qquad (۲)$$

میتوان ثابت کرد که

$$P\{G \text{ is connected}\} \approx e^{-(1-p)\sum_{i\in V} p^{k_i}} \quad (۳)$$

و در صورتی که G یک گراف k-منتظم باشد، معادله (۲) را میتوان به صورت زیر بازنویسی کرد

$$P\{G \text{ is connected}\} \approx e^{-n(1-p)p^{k}} \quad (۴)$$

از سویی، فرمولی به نام همبندی ماندهای (residual connectivity) وجود دارد که عبارتست از

$$\Phi(G) = \sum_{i=0}^{n} C_i p^i (1-p)^{n-i} \quad (۵)$$

که در آن n تعداد گرهها و p احتمال مستقل خرابیهایی از نوع گره است. Ci را تعداد مجموعه برش (cut-set) میگویند و عبارت از تعداد زیرگرافهای همبندی است که دقیقاً i گره دارد و حذف آنها سبب ناهمبندی گراف G خواهد شد. بدین ترتیب قابلیت اطمینان یا تابآوری گراف G را میتوان به صورت زیر بیان کرد

$$\mathrm{Rel}(G) = 1 - \Phi(G) \quad (۶)$$

بایستی توجه داشت که برای گرافهایی که تعداد گرهها در آنها زیاد است عملا محاسبه دقیق معادله (۶) از نوع محاسبات NP است و بنابراین برای برآورد تقریبی آن از روشهای محاسبات عددی یا شبیهسازی استفاده میشود.

در ادامه به تعاریفی اشاره خواهیم داشت و سپس خواستههای شبیهسازی را مطرح خواهیم ساخت.

یکی دیگر از موضوعات جالب و حائز اهمیت در شبکه، یافتن ارتباط بین ماتریسهای تصادفیِ یکپارچه و تحویل ناپذیر (irreducible) با مبحث تابآوری گراف است. فرض کنید برای گراف G یک ماتریس مجاورت مربعی A_{n*n} داشته باشیم که در آن عناصر واقع در غیر قطر اصلی میتوانند ۰ یا ۱ یا به شکل یک احتمال باشند. بدین ترتیب، هریک از درایههای غیر از قطر اصلی این ماتریس یک متغیر تصادفی بوده و این ماتریس تصادفی میتواند متناظر با گراف G باشد. اکنون میخواهیم ببینیم که آیا میتوان سطرها و ستونهای این ماتریس A را به گونهای جایگشت (permutation) داد تا ماتریس به فرم کاهش یافته (reduced) تبدیل شود؟

تعریف ۵: اگر بتوان ماتریس A را به یکی از دو حالت فرم بلوکی-قطری (block-diagonal form) یا فرم بلوکی-مثلثی (block-triangular form) تبدیل کرد، در اینصورت گوییم ماتریس A کاهش یافته است. طبق تعریف، ماتریس بلوکی-قطری یک ماتریس مربعی است که در آن عناصر واقع بر قطر اصلی خود ماتریسهای مربعی به اندازه دلخواه و ممکن هستند و سایر عناصر غیر واقع بر قطر اصلی دارای مقدار صفر هستند. همچنین، ماتریس بلوکی-مثلثی به فرم مثلثی بالا یا پایین است که خود عناصر ماتریس مربعی هستند که دیگر قابل تبدیل نبوده اما مابقی عناصر ماتریس مقدار صفر دارند. لازم به ذکر است که برای بررسی این نوع ماتریسها توابع و برنامههای لازم در کتابخانههای پایتون وجود دارد که با کمی جستجو میتوان آنها را به آسانی پیدا کرد. لهذا، در صورتیکه ماتریس A منتسب به گراف G نتواند به یکی از دو فرم فوق تبدیل شود، آنگاه کاهش ناپذیر تلقی خواهد شد و ما ویژگی کاهش ناپذیری را معادل با نَبود یا غیبت مولفههای ناهمبند (عدم تکه تکه شدن گراف) و در نتیجه هم ارز با استحکام و تابآوری آن در نظر خواهیم گرفت.

۳-۱-۱. روش پیشنهادی مونت-کارلو برای شبیهسازی تشخیص کاهش ناپذیری یک ماتریس منتسب به گراف

فرض کنید احتمال خرابی p معلوم باشد ($0\le p\le 1$). آنگاه میتوان به کمک آن یک ماتریس مربعی n*n ساخت. روش کار بدین ترتیب است که یک عدد تصادفی را تولید کرده و بسته به اینکه این عدد بزرگتر از یا کوچکتر از p باشد، به هر درایه از این ماتریس ۰ یا ۱ را نسبت دهید. گام بعدی این است که از ماتریس A، یک ماتریس $B=(b_{rs})_{n*n}$ را که ماتریس مسیر نام دارد، استخراج کنیم. طرز ساخت ماتریس مسیر اینگونه است که اگر بین r و s مسیری در گراف G وجود داشته باشد، درایه نظیر را ۱ و در غیر اینصورت ۰ فرض میکنیم. ماتریس B مشابه با کهاد دترمینان A است. پس از ساخت B کافیست آن را بررسی کنیم. چنانچه تمامی عناصر ماتریس B برابر ۱ باشند، یعنی همگی با هم در ارتباط هستند و در نتیجه ماتریس A کاهش ناپذیر است؛ در غیراینصورت ماتریس A کاهش پذیر خواهد بود.

خواسته ۱: برای گرافهایی از نوع تصادفی(اردوش-رینی)، دنیای کوچک (WS) و باراباشی-آلبرت (BA) که در بسته نرم افزاری NetworkX از قبل پیادهسازی شدهاند، با فرض داشتن n گره و برای مقادیر مختلف p، شبیهسازیها را به تعداد کافی برای مثال به تعداد ۱۰۰۰ بار انجام داده و متوسطگیری کنید و احتمالات کاهش ناپذیری را محاسبه و نمودارهای آنرا برحسب p ترسیم کنید. لازم است تا فاصله اطمینان و خطای شبیه سازی مونت-کارلو محاسبه و گزارش شود.

خواسته ۲: فرمولهای تحلیلی (۳) و (۶) را پیاده‌سازی کنید و نتیجه را با نتایج حاصل از شبیه‌سازی در خواسته ۱ در یک نمودار با هم نشان دهید و مقایسه کنید؟ آیا نتایج شبیه سازی شما با مدلهای تحلیلی اعتبارسنجی میشوند و مورد تایید هستند؟ اختلاف در چیست؟ دلیل این اختلاف ناشی از چیست؟

------------------------------- فاز دوم -------------------------------

۳-۲. بررسی تاب‌آوری شبکه تحت مدل خرابی پویای گره‌ها

همانطور که در بخش ۲ اشاره کردیم، مدل پویا میتواند خود به دو صورت مدل غیرفعال و فعال پیاده‌سازی گردد. در هر دو مدل، کاربران یا گره‌های شبکه میتوانند دارای طول عمر باشند که به طور مثال برای گره i طول عمر با متغیر تصادفی Li مشخص میشود که از یک توزیع احتمال (برای مثال نمایی یا پارتو و ...) پیروی کرده و بدین معنی است که پس از انقضای Li گره i بایستی شبکه را ترک کند یا به عبارت دیگر خراب و حذف شده و لینکهای متصل به i نیز حذف خواهند شد.

در بالا اشاره داشتیم که متغیر T بیانگر مدت زمان ماندن (بقاء) در شبکه قبل از ایزوله شدن است و نشاندهندهٔ وضعیتی است که در آن تمامی همسایه‌های یک گره در وضعیت off یا fail قرار میگیرند و E[T] به مقدار مورد انتظار یا متوسط T اشاره دارد. همچنین E[si] نیز به متوسط زمان جستجو برای یافتن همسایه مناسب در صورت از دست دادن همسایه‌های یک گره اشاره دارد؛ یعنی طول عمرگره i طوری است که از همسایه‌های خود بیشتر عمر میکند. توجه دارید که E[si] از E[Li] بسیار کوچکتر است. بدیهی است که در مدل غیرفعال، کاربران به دنبال جایگزینی نیستند؛ یعنی S>E[li] است و آنقدر در یافتن و جایگزینی همسایه مناسب تعلل می‌ورزند که عملاً جایگزینی مناسبی صورت نمیپذیرد. از نظر ریاضی مدل تحلیلی تقریبی که بتواند ارتباط این پارامترها را نشان دهد از قرار زیر است:

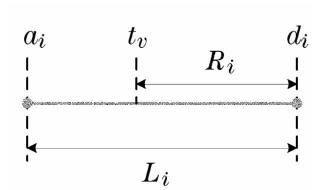
$$E[T] \approx \frac{E[s_i]}{< k >}\{(1+\frac{E[R_i]}{E[s_i]})^{<k>}-1\} \qquad (۷)$$

که در آن، <k> به معنای متوسط درجه، E[si] متوسط زمان جستجو، E[T] متوسط بقاء یا ماندن گره در شبکه قبل از ایزوله شدن و بالاخره E[Ri] متوسط زمان مانده (residual time) است.

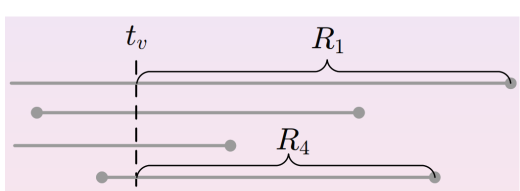
همچنین ثابت میشود که کران بالایِ احتمال ایزوله شدن شبکه از رابطه تقریبی زیر پیروی میکند

$$\varphi \leq \frac{E[L_i]}{E[T]} \qquad (۸)$$

همگی این پارامترها قبلاً شرح داده شد؛ لیکن، یک پارامتر در فرمول (۷)، یعنی E[Ri]، جدید است که در ادامه آنرا نیز شرح خواهیم داد. منظور از این پارامتر، متوسط زمان مانده برای گره i است؛ بدین معنی که اگر گره‌ای، برای مثال v، را در لحظه t در شبکه در نظر بگیریم، این گره دارای طول عمر Lv و دارای همسایگانی است که هرکدام طول عمری مثلاً Li دارند. طبق تعریف، فاصله زمانی بین لحظه ورود گره v در شبکه و لحظه خروج گره i از شبکه را (به دلیل انقضای طول عمر) زمان مانده نامیده و با Ri نشان میدهیم. برای درک بهتر به شکل زیر توجه کنید. ai یعنی arrival of node i و di یعنی departure of node i. بدین ترتیب $d_i - t_v = R_i$ و بدیهی است که $d_i - a_i = L_i$



طبیعی است که گره v میتواند همسایگانی داشته باشد که هر کدام دارای طول عمر مانده‌ای برابر R1, R2, ... هستند (به شکل زیر نگاه کنید).



خواسته ۴: تابع توزیع متغیر تصادفی طول عمر L_i را یکبار به صورت یک توزیع بی‌حافظه دُم-نازک نمایی یعنی $F(x)=1-e^{-\lambda x}$ و بار دیگر به صورت توزیع حافظه‌دار دُم-کلفت پارتو یعنی $F(x)=1-(1+x/\beta)^{-\alpha}\ ;x>0,\alpha>1$ در نظر بگیرید که در آن β>0 پارامتر مقیاس (scale) و α پارامتر شکل (shape) نام دارد. متوسط طول عمر برای توزیع پارتو عبارتست از $E[L_i]=\beta/(\alpha-1)$.

در سناریوی‌های شبیه‌سازی، مقادیر مناسب را برای هرکدام از پارامترها اختیار کنید برای مثال λ=2 در طول عمر نمایی و α=3 و β=1 برای طول عمر پارتو.

برای هر دو نوع مدل غیرفعال و فعال و نیز توزیع‌های نمایی و پارتوی طول عمر، نمودارهای زیر را ترسیم کنید. برای هر یک از نمودارها تفسیر خود را بنویسید.

الف) نمودار E[T] برحسب درجه k

ب) نمودار E[T] برحسب E[s]

ج) نمودار φ برحسب E[s]

د) نمودار φ برحسب توزیع طول عمر نمایی و پارتو

ه) تاثیر مقادیر پارامتر شکل (α) بر احتمال ایزوله شدن (φ) و متوسط زمان بقاء (E[T]) با ترسیم نمودارهای مربوطه

یادداشت ۱: *همانطور که در بالا اشاره شده این پروژه دارای دو فاز است که گرچه از منظر مفهومی با یکدیگر در ارتباط تنگاتنگ هستند، میتوانند به شکل جداگانه نیز انجام شوند. بااین حال وزن انجام فاز ۲ بیشتر است.*

یادداشت ۲: *نیازی به اثبات معادلات و روابط تحلیلی نیست و تنها از آنها برای اعتبارسنجی نتایج حاصل از شبیه‌سازی استفاده میشود. با این حال چنانچه این فرمولها اثبات شوند نمره امتیازی و ارزشمندی به اثبات آنها تعلق خواهد گرفت.*

پیروز و سرفراز باشید