**گزارش پروژه: تحلیل نفوذ در شبکه با مدل‌های Independent Cascade (IC) و Linear Threshold (LT)**

**خلاصه‌ی اجرایی**

این گزارش به صورت جامع و مرحله‌به‌مرحله پیاده‌سازی، اجرا و تحلیل آزمایش‌های نفوذپذیری (influence diffusion) روی دو شبکه‌ی تولیدشده به‌صورت نمونه Barabasi–Albert با ۱۵۰۰ گره و Watts–Strogatz با ۱۲۰۰ گره را با استفاده از دو مدل متداول انتشار

* مدل مستقل پخش (Independent Cascade — IC)
* مدل آستانه خطی (Linear Threshold — LT)

بیان می‌کند. برای هر شبکه و هر معیار مرکزیّت Degree، Closeness، Betweenness، PageRank مجموعه گره‌های برتر در اندازه‌های k = [5, 10, 20, 25, 30] استخراج شده و هر مجموعه حداقل ۱۰ بار در هر مدل شبیه‌سازی شده و میانگین نسبت گره‌های فعال‌شده به کل گره‌ها (نسبت نفوذ) گزارش گردیده است.

**مقدمه**

تحلیل انتشار اطلاعات/ویروس/نوآوری در شبکه‌ها برای درک چگونگی گسترش پدیده‌ها در سیستم‌های پیچیده اهمیت دارد. دو مدل IC و LT از مدل‌های کلاسیک در این حوزه هستند. در این پروژه هدف مقایسه کارایی معیارهای مختلف مرکزیّت در انتخاب seed set برای حداکثرسازی نفوذ در دو نوع شبکه مصنوعی متداول و مقایسه عملکرد مدل IC و LT است.

**شرح داده‌ها و شبکه‌ها**

* **شبکه اول (G1)**: Barabasi–Albert با n=1500 و m=5 (شبکه‌ای با درجه توزیع‌شده scale-free
* **شبکه دوم (G2)**: Watts–Strogatz با n=1200، k=6، p=0.1 (شبکه‌ای کوچک-جهانی با خوشه‌بندی بالا)

شبکه‌ها به‌صورت بدون‌وزن و بدون جهت (undirected) تولید شده‌اند. در صورت نیاز برای گراف‌های جهت‌دار یا وزندار می‌توان وزن یا جهت را اضافه کرد؛ ولی کد نمونه از توابع NetworkX برای تولید گراف‌های مذکور استفاده کرده است.

**معیارهای مرکزیّت و انتخاب گره‌های اولیه (Seeds)**

برای هر گراف محاسبه شد:

* Degree Centrality
* Closeness Centrality
* Betweenness Centrality
* PageRank

برای هر معیار، گره‌ها بر اساس مقدار مرکزیّت نزولی مرتب شده و k گره برتر به‌عنوان seed set انتخاب شد. مقادیر k مورد بررسی: 5، 10، 20، 25، 30.

**مدل‌های انتشار و پیاده‌سازی**

**Independent Cascade (IC)**

* در هر گام، هر گره که تازه فعال شده یک شانس مستقل p برای فعال‌سازی هر همسایه غیرفعال دارد.
* در پروژه مقدار پیش‌فرض p = 0.1 انتخاب شد (قابل تنظیم).
* فرایند تا وقتی ادامه دارد که هیچ گره جدیدی فعال نشود.

**Linear Threshold (LT)**

* هر گره آستانه‌ای (threshold) به‌صورت تصادفی در [0,1] دریافت می‌کند.
* اگر نسبت همسایگان فعال به تمام همسایگان گره از آستانه گره بیشتر یا مساوی شود، گره فعال می‌شود.
* فرایند تا همگرا شدن تکرار می‌شود.

**شبیه‌سازی‌ها**

* برای هر مجموعه seed و هر ترکیب پارامتری حداقل trials = 10 بار شبیه‌سازی انجام شد.
* خروجی هر اجرا تعداد گره‌های فعال‌شده (active) بود؛ برای هر مجموعه میانگین این تعداد بر trials محاسبه و سپس بر تعداد کل گره‌ها تقسیم شد تا نسبت نفوذ (fraction) به‌دست آید.

**جزئیات پیاده‌سازی (مروری بر کد)**

کد اصلی با استفاده از NetworkX و توابعی که خودتان نوشته‌اید اجرا شده است. مراحل کلی:

1. تولید گراف‌ها (Barabasi–Albert و Watts–Strogatz)
2. محاسبه چهار نوع مرکزیّت
3. انتخاب گره‌های top-k
4. اجرای شبیه‌سازی IC یا LT برای هر seed set به تعداد trials دفعات
5. محاسبه میانگین نفوذ و نرمال‌سازی بر n
6. ذخیره نتایج در DataFrame و نمایش نموداری

پارامترهای کلیدی که در پروژه استفاده شد:

* ks = [5, 10, 20, 25, 30]
* trials = 10
* p (for IC) = 0.1
* اندازه گراف 1500 و 1200

**خروجی‌ها و نمودارها (محل قرار‌گیری تصاویر و جداول)**

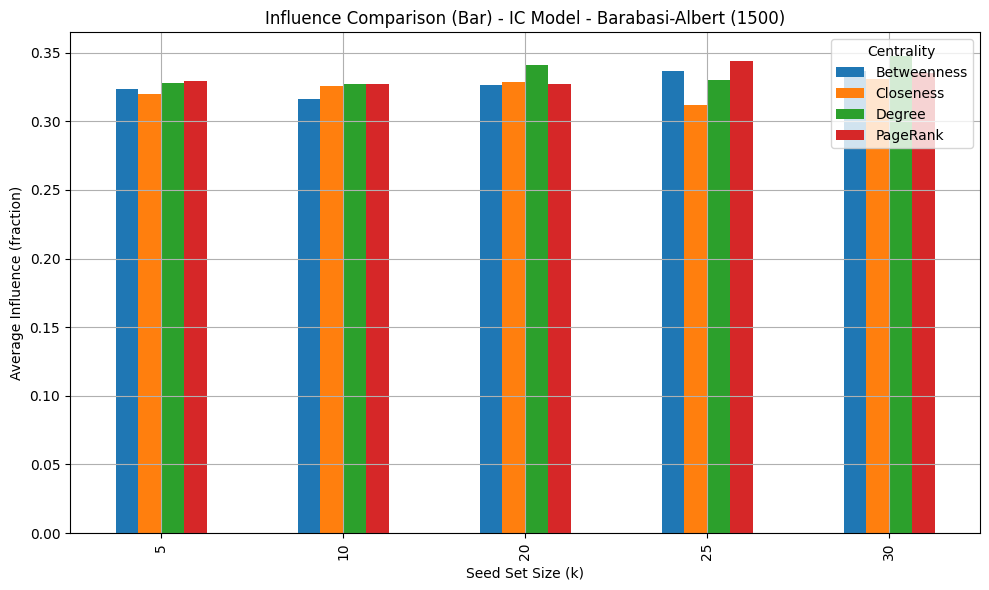
در ادامه مکان‌هایی برای تصاویر (نمودارها) و جداول نتایج قرار داده شده است. فایل‌های تصویری و جداول وقتی از اجرای کد روی داده‌های واقعی حاصل شوند باید در این محل‌ها قرار گیرند. نام‌گذاری پیشنهادی برای فایل‌های تصویری و جداول هم آورده شده است.

**نتایج مدل IC**

**توضیح کلی:** برای هر شبکه دو نوع نمایش برای نتایج تولید می‌شود: (الف) نمودار خطیِ سهم نفوذ (fraction) در برابر اندازه seed set (k) برای هر معیار مرکزیّت؛ (ب) نمودار ستونی (bar) مقایسه‌ای که برای هر k مقدار نفوذ هر معیار را کنار هم نشان می‌دهد. همچنین دو جدول pivot که مقادیر عددی میانگینِ نفوذ برای هر ترکیب (Centrality × k) نمایش می‌دهند.

**عکس‌ها (IC)**

1. **شکل IC-1:** نمودار خطی (Line plot) — شبکه: Barabasi–Albert (1500) — مقایسه 4 مرکزیّت و ... بر حسب k  
   *A graph with lines and numbers

   AI-generated content may be incorrect.*
2. **شکل IC-2:** نمودار ستونی (Bar chart) — شبکه: Barabasi–Albert (1500) — pivot over k.  
   **
3. **شکل IC-3:** نمودار خطی — شبکه: Watts–Strogatz (1200) — مقایسه 4 مرکزیّت و .... بر حسب k  
   *A graph with different colored lines

   AI-generated content may be incorrect.***شکل IC-4:** نمودار ستونی — شبکه: Watts–Strogatz (1200) — pivot over k.  
   *A graph of different colored bars

   AI-generated content may be incorrect.*

**جدول‌ها (IC)**

* **جدول IC-1 (Barabasi–Albert 1500):** جدول pivot با index = k و ستون‌ها = Degree, Closeness, Betweenness, PageRank و مقادیر = Influence (fraction).  
  *A screenshot of a computer screen

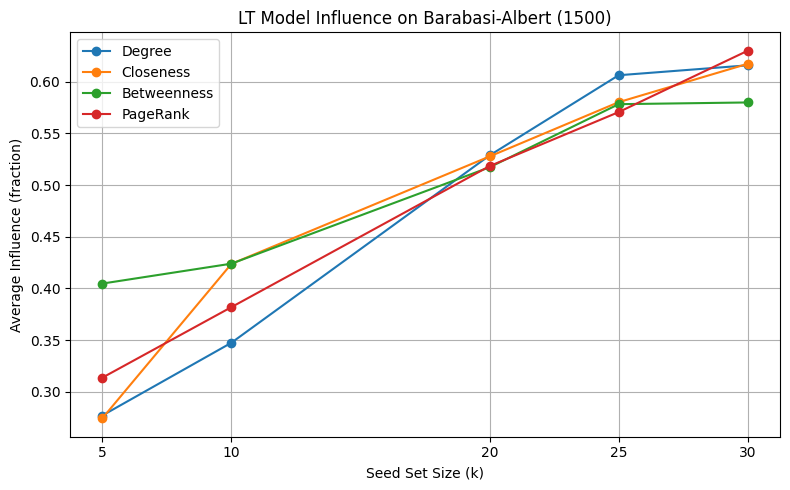
  AI-generated content may be incorrect.***جدول IC-2 (Watts–Strogatz 1200):** همانند بالا برای شبکه دوم.  
  *A screenshot of a computer screen

  AI-generated content may be incorrect.*

**نتایج مدل LT**

**توضیح کلی:** ساختار نتایج مانند بخش IC است؛ نمودارها و جداول جداگانه برای هر شبکه.

**عکس‌ها (LT)**

1. **شکل LT-1:** نمودار خطی — شبکه: Barabasi–Albert (1500) — مقایسه 4 مرکزیّت و .... بر حسب k   
   
2. **شکل LT-2:** نمودار ستونی — شبکه: Barabasi–Albert (1500).  
   *A graph of different colored bars

   AI-generated content may be incorrect.*
3. **شکل LT-3:** نمودار خطی — شبکه: Watts–Strogatz (1200).  
   *A graph with different colored lines

   AI-generated content may be incorrect.*
4. **شکل LT-4:** نمودار ستونی — شبکه: Watts–Strogatz (1200).  
   *A graph of different colored bars

   AI-generated content may be incorrect.*

**جدول‌ها (LT)**

* **جدول LT-1 (Barabasi–Albert 1500):** pivot table مشابه برای LT.  
  *A screenshot of a computer screen

  AI-generated content may be incorrect.*
* **جدول LT-2 (Watts–Strogatz 1200):** pivot table مشابه برای شبکه دوم.  
  *A screen shot of a black screen

  AI-generated content may be incorrect.*

**تصاویر مقایسه‌ای IC vs LT**

برای نشان‌دادن تفاوت الگوهای نفوذ بین دو مدل دو تصویر مقایسه‌ای پیشنهاد می‌شود:

1. **شکل COMP-1:** نمودار ترکیبی (برای هر مرکزیّت، دو منحنی IC و LT) — شبکه: Barabasi–Albert (1500).  
   *A graph of different colored lines

   AI-generated content may be incorrect.*
2. **شکل COMP-2:** نمودار ترکیبی — شبکه: Watts–Strogatz (1200).  
   *A graph of different colored lines

   AI-generated content may be incorrect.*

**تفسیر نتایج**

* **افزایش k معمولاً نفوذ را افزایش می‌دهد، اما نرخ بازدهی کاهشی است.** یعنی از k=5 تا k=30 نفوذ بیشتر می‌شود اما هر واحد اضافه تأثیر کمتری دارد.
* **تفاوت معیارها بسته به ساختار شبکه متغیر است.** در شبکه‌های scale-free (مثل BA) معیارهای مبتنی بر درجه یا PageRank معمولاً نتایج خوبی دارند، چون چند گره با درجه بسیار بالا وجود دارد؛ اما در شبکه‌های با خوشه‌بندی بالا (مثل WS) معیارهایی که موقعیت پل‌مانند (bridge) را شناسایی می‌کنند (مثلاً Betweenness) ممکن است عملکرد بهتری در گسترش از خوشه‌ای به خوشه دیگر داشته باشند.
* **تفاوت IC و LT:** IC بر انتقال تصادفیِ لبه‌به‌لبه مبتنی است در حالی که LT نیازمند تجمع فشار از همسایگان است؛ بنابراین مرکزیّت‌هایی که گره‌هایی با اتصالات زیاد به خوشه‌های مختلف انتخاب می‌کنند در IC سریع‌تر نفوذ می‌کنند؛ در LT، توزیع همسایگی و آستانه‌ها نقش پررنگ‌تری دارد.