



دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

رساله دکتری
مهندسی کامپیوتر

انتخاب مدل برای شبیه‌سازی شبکه‌های اجتماعی در یک سیستم تصمیم‌یار

نگارش
صادق علی‌اکبری

استاد راهنما
دکتر جعفر حبیبی و دکتر علی موقر

۲۰ اسفند ۱۳۹۳

به نام او

دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

رساله دکتری

عنوان: انتخاب مدل برای شبیه‌سازی شبکه‌های اجتماعی در یک سیستم تصمیم‌یار
نگارش: صادق علی‌اکبری

کمیته داوران

استاد راهنما: دکتر جعفر حبیبی و دکتر علی موقر امضاء:

استاد ممتحن داخلی: دکتر مهدی جلیلی امضاء:

استاد ممتحن داخلی: دکتر محمدعلی آبام امضاء:

استاد ممتحن خارجی: دکتر احمد خونساری امضاء:

استاد ممتحن خارجی: دکتر مسعود اسدیپور امضاء:

استاد ممتحن خارجی: دکتر مصطفی صالحی امضاء:

تاریخ:

تقدیم به پدر عزیزم،

که اولین معلم، و بهترین معلم من بوده و هست،

که به من الفبا و اعداد را آموخت و الفبای زندگی را ...

سپاس نامه

در ابتدا، از خانواده عزیزم بابت مهربانی‌ها، حمایت‌ها و دلگرمی‌های بی‌دریغشان ممنونم. دوره دکتری در ایران دوره سختی است و تکالیف دانشجوی دکتری با حقوق و اختیارات وی هم‌خوان و متوازن نیست، و این واقعیت حتماً باعث شده که در این مدت، نقشی که به عنوان عضوی از خانواده به عهده داشته‌ام، به شایستگی ادا نکنم. از این جهت، از همسر مهربانم و پسر عزیزم بابت همراهی و صبرشان ممنونم. همچنین از پدر و مادر عزیزم و خواهر و برادر خوبم، بابت حمایت‌ها و دلگرمی‌هایشان سپاسگزارم. از اهالی خانه سبزمان نیز بابت مهربانی‌ها و دلسوزی‌های بی‌پایانشان ممنونم.

از اساتید محترم راهنما و ممتحن که دلسوزانه در به ثمر رسیدن این رساله کوشیدند، سپاسگزارم. همچنین در دوره دکتری، با برخی از دوستان، افتخار همکاری در کارهای پژوهشی داشتم و به‌ویژه از صادق مُطَلّبی، سینا رشیدیان و جواد قره‌چمنی، بابت انرژی مثبت و تلاش مؤثرشان ممنونم. همچنین قدردان اساتید و دوستانی هستم که صمیمانه و دلسوزانه به بنده مشورت دادند، به‌ویژه دکتر حسین رحمانی، دکتر عباس حیدرنوری، دکتر مهدیه سلیمانی، دکتر مهدی جلیلی و دکتر مسعود اسدپور. با وجود همه مشکلات و سختی‌ها، کار و پژوهش در فضای صمیمی دانشجویان و دوستانی هم‌چون محمدحسین زنگویی، شهروز معاون، محمد تنهایی، محمدمامین فضلی و محمود نشاطی بسیار دلگرم‌کننده و امیدبخش بود. همچنین از مدیران محترم شرکت مشاوران نرم‌افزاری اعوان که از نظر علمی، فنی و معنوی پشتیبان من بودند، تقدیر می‌نمایم.

این پروژه تحت قرارداد پژوهشی به شماره ۵۰۰/۱۱۳۰۷/ت مورخ ۹۱/۷/۲۲ از حمایت و پشتیبانی مالی و معنوی پژوهشگاه ارتباطات و فناوری اطلاعات (مرکز تحقیقات مخابرات ایران) بهره‌مند شده است.

وز هر چه گفته اند و شنیدیم و خوانده ایم
ما همچنان در اوّل وصف تو مانده‌ایم...

ای برتر از خیال و قیاس و گمان و وهم
مجلس تمام گشت و به آخر رسید عمر

چکیده

یک شبکه اجتماعی، مجموعه‌ای از موجودیت‌ها و روابط بین آن‌ها را بازنمایی می‌کند. روابط اجتماعی روزمره افراد جامعه، شبکه ارتباطات تلفنی، شبکه‌های اجتماعی مجازی و شبکه نویسندگان مقالات، نمونه‌هایی از شبکه‌های اجتماعی هستند. با گسترش روزافزون شبکه‌های اجتماعی، بر اهمیت آن‌ها افزوده شده است و تحلیل این شبکه‌ها کاربردهای متنوعی یافته است. به‌ویژه، مدیران و تصمیم‌سازان چنین شبکه‌هایی نیازمند سیستم‌هایی هوشمند برای پشتیبانی تصمیم در جهت آینده‌پژوهی و مدیریت این شبکه‌ها هستند. این سیستم‌های تصمیم‌یار، امکان تعریف پارامترهای مسأله اجتماعی موردنظر و تحلیل "چه می‌شود اگر" را برای تصمیم‌سازان فراهم می‌سازد. یکی از روش‌های موجود در این زمینه، شبیه‌سازی رایانه‌ای است. در این رویکرد، می‌توانیم شبکه موردنظر را شبیه‌سازی کنیم و سناریوهای مختلف را روی آن بیازماییم و نتایج را بررسی کنیم. شبیه‌سازی، روشی مؤثر برای بررسی نظریه‌ها و یا طراحی نظریه‌های جدید در حوزه‌های مختلف اجتماعی است. هدف این رساله پژوهشی، ارائه راهکارهایی برای برخی خلأهای پژوهشی برای رسیدن به یک سیستم تصمیم‌یار مبتنی بر شبیه‌سازی شبکه‌های اجتماعی است. در سیستم موردنظر، مدیر (تصمیم‌ساز) می‌تواند ویژگی‌های شبکه اجتماعی موردنظر را تعیین کند، یک شبکه مصنوعی با اندازه و خواص موردنظر تولید کند، سناریوها و فرایندهای مختلف را در این شبکه به اجرا بگذارد و نتیجه شبیه‌سازی را مشاهده و بررسی کند.

یکی از مسایل مهم این پژوهش، تولید شبکه‌های اجتماعی مصنوعی، به منظور استفاده در فرایند شبیه‌سازی است. شبکه‌های مصنوعی، گراف‌هایی تصادفی هستند که از نظر ساختاری، شبیه شبکه موردنظر (شبکه هدف) هستند. در میان مدل‌های تولید شبکه مصنوعی، روش جامع و مانعی که برای همه کاربردها و همه شبکه‌های هدف مناسب باشد وجود ندارد. در این رساله، چارچوبی برای تولید شبکه‌های مصنوعی ارائه کردیم که بر مبنای حل چند زیرمسأله استوار است. زیرمسأله اول، استخراج ویژگی‌های ساختاری مهم از شبکه‌های اجتماعی است. در این بخش، روشی کمی‌سازی توزیع درجه در شبکه‌های اجتماعی ارائه کردیم، که دقیق‌تر از روش‌های موجود است و وابسته به اندازه شبکه نیست. زیرمسأله دوم، نیاز به یک معیار شباهت بین ساختارهای شبکه‌های اجتماعی است. در این بخش، روشی ارائه کردیم که دقیق‌تر و کاراتر از روش‌های موجود است. زیرمسأله سوم، انتخاب مدل مولد مناسب برای تولید شبکه‌های اجتماعی است. در این بخش دو روش مختلف پیشنهاد کرده‌ایم که هر دو دقیق‌تر از روش‌های موجود هستند. زیرمسأله چهارم، تنظیم مدل انتخاب شده با کمک پارامترهای مناسب برای تولید شبکه‌هایی مشابه شبکه هدف است. در این زیرمسأله نیز روشی دقیق و مقاوم نسبت به نویز ارائه کرده‌ایم. با توجه به پیوستگی زیرمسأله‌های مورد اشاره، این رساله یکی از اولین و گسترده‌ترین پژوهش‌ها در زمینه انتخاب، تنظیم و ارزیابی هوشمند مدل‌های مولد شبکه‌های اجتماعی محسوب می‌شود. در نهایت، ارائه مدل و معماری نرم‌افزاری برای شبیه‌سازی شبکه‌های بزرگ، نیازمندی مهمی است که در این مسأله نیز به ارائه یک مدل مرجع اولیه نرم‌افزاری پرداخته‌ایم.

کلمات کلیدی: شبکه اجتماعی، شبیه‌سازی اجتماعی، سیستم تصمیم‌یار، مدل مولد شبکه، داده‌کاوی، شبکه پیچیده، سیستم چند عامله

فهرست مطالب

چکیده

۱	مقدمه	۱
۱	شبکه‌های اجتماعی	۱.۱
۲	شبیه‌سازی اجتماعی	۲.۱
۲	کاربردهای شبیه‌سازی شبکه‌های اجتماعی	۳.۱
۳	سیستم تصمیم‌یار مبتنی بر شبیه‌سازی اجتماعی	۴.۱
۴	مسائل اساسی تحقیق	۵.۱
۵	ساختار رساله	۶.۱
۷	مرور ادبیات موضوع	۲
۷	ویژگی‌های ساختاری معمول در شبکه‌های اجتماعی	۱.۲
۷	گراف تُنک	۱.۱.۲
۷	دنیای کوچک	۲.۱.۲
۸	توزیع درجه گره‌ها	۳.۱.۲
۹	ساختار گروهی شبکه	۴.۱.۲
۱۱	ویژگی تراگذری یال‌ها	۵.۱.۲
۱۱	هم‌رنگی	۶.۱.۲
۱۱	مدل‌های تولید شبکه اجتماعی تصادفی	۲.۲
۱۵	مجموعه داده‌ها	۳
۱۵	شبکه‌های حقیقی	۱.۳
۱۷	شبکه‌های مصنوعی	۲.۳
۱۸	شبکه‌های زمان‌دار	۳.۳

۲۱	استخراج ویژگی از توزیع درجه گره‌های شبکه	۴
۲۴	۱.۴ مرور روش‌های کمی‌سازی و مقایسه توزیع درجه	
۲۵	۲.۴ روش پیشنهادی برای کمی‌سازی و مقایسه توزیع درجه	
۲۷	۳.۴ ارزیابی	
۲۸	۱.۳.۴ معیارهای ارزیابی	
۲۹	۲.۳.۴ نتایج ارزیابی	
۳۹	۴.۴ جمع‌بندی	
۴۱	ارائه معیار شباهت برای شبکه‌های اجتماعی	۵
۴۳	۱.۵ مرور ادبیات	
۴۵	۲.۵ یادگیری تابع فاصله برای شبکه‌های اجتماعی	
۴۶	۱.۲.۵ شواهد شباهت شبکه‌ها	
۴۶	۲.۲.۵ استخراج ویژگی	
۴۸	۳.۲.۵ انتخاب ویژگی و یادگیری تابع فاصله	
۴۹	۳.۵ ارزیابی	
۴۹	۱.۳.۵ مجموعه داده‌ها	
۵۰	۲.۳.۵ تابع NetDistance	
۵۱	۳.۳.۵ کارآمدی یادگیری ماشین برای ایجاد تابع فاصله	
۵۱	۴.۳.۵ روش‌های پایه	
۵۴	۵.۳.۵ آزمایش‌ها	
۶۱	۴.۵ پیچیدگی محاسباتی و حافظه‌ای	
۶۳	۵.۵ جمع‌بندی	
۶۴	۶.۵ پیوست‌ها	
۶۴	۱.۶.۵ پیوست الف: الگوریتم ژنتیک	
۶۴	۲.۶.۵ پیوست ب: جزئیات پیاده‌سازی	
۶۷	انتخاب و تنظیم مدل برای تولید شبکه اجتماعی مصنوعی	۶
۶۸	۱.۶ مرور ادبیات	
۶۹	۲.۶ انتخاب مدل: روش ModelFit	
۷۱	۳.۶ ارزیابی روش انتخاب مدل	
۷۱	۱.۳.۶ مجموعه داده‌ها	
۷۲	۲.۳.۶ روش‌های پایه	

۷۳	نتایج ارزیابی	۳.۳.۶
۷۶	تخمین پارامترهای مدل در روش ModelFit	۴.۶
۸۰	جمع‌بندی	۵.۶
۸۰	پیوست الف: انتخاب مدل در روش GMSCN	۶.۶
۸۳	۷ نتیجه‌گیری	
۸۳	جمع‌بندی	۱.۷
۸۴	خلاصه نتایج	۲.۷
۸۵	کارهای آینده	۳.۷
۹۷	کتاب‌نامه	
۱۱۱	واژه‌نامه انگلیسی به فارسی	
۱۱۵	واژه‌نامه فارسی به انگلیسی	

لیست تصاویر

۹	توزیع درجه نمایی در شبکه‌های مستقل از مقیاس	۱.۲
۱۰	توزیع توانی درجه رئوس در گراف صفحات وب [۱]	۲.۲
۱۰	شبکه‌ای با ساختار گروهی	۳.۲
۱۲	نمونه گراف‌های تصادفی که با مدل ER ساخته شده‌اند [۲]	۴.۲
۱۳	گراف‌های ایجاد شده توسط مدل دنیای کوچک Watts–Strogatz [۲]	۵.۲
	توزیع درجه چهار شبکه اجتماعی که همگی دم‌دراز هستند. توزیع درجه شبکه‌های ارجاعات	۱.۴
۲۳	به هم شبیه‌تر هستند و شبکه‌های همکاری نیز به یکدیگر شبیه‌ترند.	۲.۴
	توزیع درجه چهار شبکه ارجاعات در گذر زمان. هر چه شبکه‌ها بزرگتر می‌شوند، بیشینه،	۲.۴
۲۴	میانگین و انحراف از معیار درجه گره‌ها هم افزایش می‌یابد.	۳.۴
	میانگین دقت KNN در شبکه‌های مصنوعی (برای $K = ۱۰$). در این مجموعه داده،	۳.۴
	دقت روش پیشنهادی (DDQC)، همه روش‌های پایه را با اختلافی بیش از ۲۰ درصد	۳.۰
	پشت سر می‌گذارد.	۴.۴
	میانگین دقت KNN در شبکه‌های حقیقی (برای $K = ۱۰$). در این مجموعه داده نیز	۳.۱
	روش پیشنهادی (DDQC) بهترین است.	۵.۴
	دقت KNN برای شبکه‌های مصنوعی به ازای مقادیر مختلف K . در این مجموعه داده،	۳.۱
	روش پیشنهادی (DDQC) از همه روش‌های پایه بهتر عمل می‌کند.	۶.۴
	دقت KNN برای شبکه‌های حقیقی به ازای مقادیر مختلف K . در این مجموعه داده، روش	۳.۲
	پیشنهادی (DDQC) در اکثر مقادیر K از روش‌های پایه بهتر عمل می‌کند.	۷.۴
	دقت KNN در مجموعه شبکه‌های مصنوعی به ازای مدل‌های مولد مختلف. در بیشتر	۳.۲
	مدل‌ها DDQC بهتر عمل می‌کند.	۸.۴
	دقت KNN در مجموعه شبکه‌های حقیقی به ازای انواع مختلف شبکه. در بیشتر انواع	۳.۳
	شبکه، DDQC بهتر عمل می‌کند.	۹.۴
	میانگین $P@K$ در شبکه‌های مصنوعی برای $(K = ۱۰)$. بر اساس این معیار ارزیابی،	۳.۳
	روش پیشنهادی بیش از ۱۸ درصد از همه روش‌های پایه پیش افتاده است.	

۱۰.۴	میانگین P@K در شبکه‌های حقیقی برای ($K = ۱۰$). روش پیشنهادی نتایجی بهتر از
۳۴	روش‌های پایه نشان می‌دهد.
۱۱.۴	شاخص Dunn در شبکه‌های مصنوعی. از منظر شاخص Dunn، روش پیشنهادی بیش
۳۴	از ۱۶ درصد بهتر از روش‌های پایه عمل می‌کند.
۱۲.۴	شاخص Dunn در شبکه‌های حقیقی. در این مجموعه داده هم روش پیشنهادی حداقل ۱۶
۳۵	درصد بهتر از روش‌های پایه از منظر شاخص Dunn عمل می‌کند.
۱۳.۴	پایداری روش‌های مختلف نسبت به رشد اندازه شبکه در مجموعه شبکه‌های مصنوعی. در
	مجموعه داده‌های مصنوعی روش‌های DDQC و Percentiles برای شبکه‌های بزرگ‌تر
	پایدارتر می‌شوند. به خصوص روش پیشنهادی (DDQC) پایداری بهتری را در شبکه‌هایی
۳۶	با بیش از ۱۰۰۰ گره نشان می‌دهد.
۱۴.۴	پایداری روش‌های مختلف نسبت به رشد اندازه شبکه در مجموعه شبکه‌های حقیقی (محور
	افقی با نمای ۱۰ بزرگ می‌شود). در مجموعه داده‌های حقیقی روش‌های DDQC و
۳۷	Percentiles برای شبکه‌های بزرگ‌تر پایدارتر می‌شوند.
۱۵.۴	رده‌بندی شبکه‌های مصنوعی براساس روش SVM. روش پیشنهادی (DDQC)
۳۸	ویژگی‌هایی را به بردار ویژگی اضافه می‌کنند که بهترین رده‌بندی را نتیجه می‌دهند.
۱۶.۴	رده‌بندی شبکه‌های حقیقی براساس روش SVM. روش پیشنهادی (DDQC) ویژگی‌هایی
۳۹	را به بردار ویژگی اضافه می‌کنند که بهترین رده‌بندی را نتیجه می‌دهند.
۱.۵	مسیر یادگیری و ارزیابی تابع فاصله برای شبکه‌های اجتماعی.
۲.۵	میانگین دقت رده‌بند KNN (به ازای $K = ۱۰$) بر مبنای روش‌های مختلف مقایسه
۵۲	(توابع فاصله مختلف) در مجموعه داده شبکه‌های مصنوعی.
۳.۵	میانگین دقت رده‌بند KNN (به ازای $K = ۱۰$) بر مبنای روش‌های مختلف مقایسه در
۵۲	مجموعه داده شبکه‌های حقیقی.
۴.۵	دقت رده‌بند KNN به ازای مقادیر مختلف K در مجموعه شبکه‌های مصنوعی براساس
۵۴	توابع مختلف فاصله.
۵.۵	دقت رده‌بند KNN به ازای مقادیر مختلف K در مجموعه شبکه‌های حقیقی براساس توابع
۵۵	مختلف فاصله.
۶.۵	معیار P@K به ازای مقادیر مختلف K در مجموعه شبکه‌های مصنوعی براساس توابع فاصله
۵۵	مختلف.
۷.۵	معیار P@K به ازای مقادیر مختلف K در مجموعه شبکه‌های حقیقی براساس توابع فاصله
۵۵	مختلف.

۵۶	۸.۵	نسبت میانگین فاصله شبکه‌های ناهم‌نوع به میانگین فاصله شبکه‌های هم‌نوع در مجموعه شبکه‌های مصنوعی.
۵۷	۹.۵	نسبت میانگین فاصله شبکه‌های ناهم‌نوع به میانگین فاصله شبکه‌های هم‌نوع در مجموعه شبکه‌های حقیقی.
۵۷	۱۰.۵	شاخص Dunn برای توابع فاصله مختلف در مجموعه شبکه‌های مصنوعی.
۵۸	۱۱.۵	شاخص Dunn برای توابع فاصله مختلف در مجموعه شبکه‌های حقیقی.
۵۹	۱۲.۵	میزان همبستگی بین فاصله ساختاری شبکه‌ها و فاصله زمانی آن‌ها در مجموعه شبکه‌های ارجاعات Cit_CiteSeerX.
۶۰	۱۳.۵	میزان همبستگی بین فاصله ساختاری شبکه‌ها و فاصله زمانی آن‌ها در مجموعه شبکه‌های همکاری Collab_CiteSeerX.
۶۰	۱۴.۵	فاصله هر شبکه ارجاعات با برهه‌های همسایه زمانی آن در گذر زمان که بر بیشینه فواصل تقسیم شده‌اند.
۶۱	۱۵.۵	فاصله هر شبکه همکاری‌ها با برهه‌های همسایه زمانی آن در گذر زمان که بر بیشینه فواصل تقسیم شده‌اند.
۶۲	۱۶.۵	دقت رده‌بند KNN برای مجموعه شبکه‌های مصنوعی به ازای مقادیر مختلف K در دو حالت: ۱- مقایسه شبکه‌هایی با اندازه مشابه ۲- مقایسه همه شبکه‌ها.
۶۵	۱۷.۵	مفاهیم و عملگرهای مهم در الگوریتم ژنتیک.
۷۰	۱.۶	مسیر پیشنهادی برای انتخاب مدل در روش ModelFit.
۷۴	۲.۶	دقت روش‌های مختلف انتخاب مدل.
۷۵	۳.۶	تأثیر نویز در روش‌های مختلف انتخاب مدل.
۷۵	۴.۶	تأثیر اندازه مجموعه همسایه‌ها بر دقت رده‌بندی ModelFit.
۷۹	۵.۶	ضریب همبستگی پیرسون بین پارامترهای تخمینی و پارامترهای واقعی در مدل‌های مولد مختلف.
۸۵	۱.۷	مهمترین دستاوردها و نوآوری‌های ارائه شده در این رساله.
۹۱	۲.۷	ارتباط اجزای پایگاه داده پیش‌نگر
۹۱	۳.۷	نحوه تعامل عامل‌ها در محیط شبیه‌سازی

لیست جداول

۱۶	مجموعه شبکه‌های حقیقی	۱.۳
۴۸	ویژگی‌های ساختاری استخراج شده از هر شبکه	۱.۵
	وزن ویژگی‌های انتخاب شده در معیار فاصله NetDistance که در قالب تابع فاصله منهن	۲.۵
۵۰	وزن‌دار پیشنهاد شده است.	
۶۲	پیچیدگی زمانی و حافظه‌ای مؤلفه‌های استخراج ویژگی در تابع NetDistance	۳.۵
۶۳	پیچیدگی زمانی و حافظه‌ای روش‌های مختلف محاسبه فاصله شبکه‌ها.	۴.۵
۷۸	میانگین دقت تخمین پارامتر در روش پیشنهادی و روش مبتنی بر فاصله	۱.۶
۷۹	خطای میانگین مربعات (MSE) برای پارامترهای تخمینی.	۲.۶
	ضریب همبستگی پیرسون (R) بین پارامترهای تخمینی و پارامترهای واقعی در روش	۳.۶
۷۹	پیشنهادی (ModelFit) به ازای مقادیر مختلف نویز.	
	مقایسه دقت روش ModelFit و روش KronFit برای تخمین پارامترهای مولد در	۴.۶
۸۰	شبکه‌هایی که با مدل Kronecker تولید شده‌اند.	
۸۹	گرامر زبان پرسمان برای پایگاه داده پیش‌نگر	۱.۷

۱ مقدمه

۱.۱ شبکه‌های اجتماعی

یک شبکه اجتماعی مجموعه‌ای از عناصر و ارتباطات آن‌ها با یکدیگر را نشان می‌دهد. عناصر موجود در شبکه‌های اجتماعی معمولاً انسان‌ها هستند و ارتباط بین عناصر، نشان‌دهنده رابطه بین انسان‌هاست. گاهی نیز برخی دیگر از موجودیت‌ها، که ماهیت اجتماعی دارند، مانند سازمان‌ها، گروه‌ها و حتی عناصری مانند اشیاء و تلفن‌های همراه، اعضای شبکه‌های اجتماعی را تشکیل می‌دهند. ارتباط در شبکه‌های اجتماعی نیز واقعیت‌های مختلفی مانند دوستی، تعامل، دادوستد، وجود ارتباط و غیره را در دنیای واقعی تداعی می‌کنند. شبکه‌های اجتماعی معمولاً به صورت یک گراف بازنمایی می‌شوند که در آن گره‌ها عناصر و موجودیت‌های شبکه هستند و یال‌ها ارتباط بین موجودیت‌های اجتماعی را نشان می‌دهند. امروزه شبکه‌های اجتماعی از اهمیت فراوانی برخوردار هستند. شبکه‌های اجتماعی از منابعی مانند ارتباطات ایمیلی، روابط شغلی و سازمانی، مکالمات مخابراتی، و حتی اطلاعات مکانی قابل استخراج هستند. همچنین با پیدایش سرویس‌های برخط شبکه‌های اجتماعی، بسیاری از کاربران جامعه اطرافیان خود را در این سرویس‌ها آشکار می‌کنند و زمان قابل توجهی را برای تعامل در این شبکه‌ها صرف می‌کنند.

تحلیل شبکه‌های اجتماعی از موضوعات مهم در حوزه شبکه‌های اجتماعی است. هر چند این گونه از تحلیل تاریخچه طولانی دارد و ریشه آن در جامعه‌شناسی و انسان‌شناسی اجتماعی است، اما در سال‌های اخیر اهمیتی دوچندان یافته است. هدف از تحلیل شبکه‌های اجتماعی، تشخیص و پیش‌بینی ویژگی‌های گره‌ها، زیرمجموعه‌ها، ارتباطات، نحوه تغییرات آن‌ها و تحلیل فرایندهای حاکم بر شبکه است. روش‌ها و کاربردهای مختلفی برای تحلیل شبکه‌های اجتماعی وجود دارد، که از آن میان می‌توان به این موارد اشاره کرد: تشخیص گروه‌ها و خوشه‌های اجتماعی، تشخیص ارتباطات کلیدی و جایگاه‌های مهم در شبکه، تشخیص افراد پرنفوذ، پیش‌بینی تغییرات ساختاری در شبکه و الگوهای انتشار در شبکه. تحلیل شبکه‌های اجتماعی یک حوزه بین‌رشته‌ای (و در واقع فرارشته‌ای) است و در شاخه‌های مختلفی مانند جامعه‌شناسی، انسان‌شناسی، بیولوژی، ارتباطات، ریاضیات، فیزیک، روانشناسی اجتماعی، مدیریت، و علوم رایانه‌ای مورد بررسی قرار می‌گیرد. تحلیل شبکه‌های اجتماعی در سال‌های اخیر اهمیتی بیش از پیش یافته‌اند. دلایل مختلفی مانند پیدایش و نفوذ اینترنت، همه‌گیر شدن سرویس‌های برخط شبکه‌های اجتماعی، افزایش قدرت رایانه‌ها، پیشرفت‌های علمی و الگوریتمی در حوزه علوم رایانه و هوش مصنوعی و از همه مهم‌تر، کاربردهای فراوان این موضوع در علوم مختلف، بر

اهمیت علم شبکه و تحلیل شبکه‌ها افزوده است. در این رساله، از اصطلاحاتی مانند «شبکه»، «شبکه اجتماعی» و «شبکه پیچیده» با یک منظور استفاده می‌کنیم.

۲.۱ شبیه‌سازی اجتماعی

یکی از رویکردهای تحلیل شبکه‌های اجتماعی، شبیه‌سازی سناریوها و تحلیل فرایندها و پیش‌بینی آن‌ها در آینده است. بیشتر روش‌های کنونی مانند روش‌های تشخیص گروه‌ها، تعیین اهمیت و قدرت گره‌ها و یال‌ها، بررسی ویژگی‌های آماری گره‌ها و یال‌ها و خواص کلی گراف‌ها، وضعیت فعلی شبکه را تحلیل می‌کنند و ویژگی‌های آن را استخراج می‌نمایند. در نسل دیگری از روش‌های تحلیلی، یک شبکه اجتماعی با ویژگی‌های خاص خودش به همراه تعاملات ممکن در شبکه شبیه‌سازی می‌شود و مدیر شبیه‌سازی می‌تواند سناریوها و حالت‌های مختلف اجرایی را بررسی کند. چنین تحلیلی بخشی از یک سیستم تصمیم‌یار مبتنی بر شبیه‌سازی برای تحلیل شبکه‌های اجتماعی خواهد بود. منظور از شبیه‌سازی شبکه‌های اجتماعی، مدل‌سازی ویژگی‌های شبکه اجتماعی و اجرای مجازی فرایندها و تعاملات در شبکه اجتماعی است. به عنوان مثال، شبیه‌سازی انتشار شایعه (یا بیماری) در جامعه، تأثیر تبلیغات بر گروه‌های مختلف، الگوی برقراری ارتباط بین افراد و تأثیر عوامل مختلف اجتماعی در بازار، نمونه‌هایی از شبیه‌سازی شبکه‌های اجتماعی هستند.

امروزه شبیه‌سازی در علوم اجتماعی نیز بسیار مورد توجه قرار گرفته است. در گذشته دو رویکرد بر علوم اجتماعی سنتی مسلط بوده است: رویکرد استقرایی و رویکرد قیاسی. در رویکرد قیاسی از روی قواعد موجود قواعد جدید استنتاج می‌شود و در رویکرد استقرایی، با مشاهده مثال‌ها و نمونه‌های فراوان، مدل‌ها و قوانین جدید کشف می‌شوند. با پیشرفت علوم رایانه‌ای، شبیه‌سازی به عنوان سومین رویکرد علمی در این زمینه مورد توجه قرار گرفته است [۳]. شبیه‌سازی برخلاف قیاس، قضیه خاصی را اثبات نمی‌کند و برخلاف استقراء، بر روی داده‌های جهان واقعی کار نمی‌کند. اما داده‌های شبیه‌سازی برای استفاده در یک فرایند استقرایی برای استخراج الگو و یا آزمایش یک مدل برآمده از قیاس در قوانین شبیه‌سازی قابل استفاده است. شبیه‌سازی به عنوان ابزاری برای توسعه تئوری‌های اجتماعی-شناختی استفاده می‌شود [۴]. آنچه شبیه‌سازی اجتماعی را از دیگر مدل‌ها و تئوری‌های اجتماعی (مانند تئوری بازی‌ها) متمایز می‌کند، امکان تعامل بیشتر و مؤثرتر با مشاهده‌ها و داده‌ها است [۴]. به همین دلیل چنین ابزاری امکان اجرای سناریوهای «چه می‌شود اگر» را به کاربر می‌دهد که این امکان بخش مهمی از یک سیستم تصمیم‌یار است.

۳.۱ کاربردهای شبیه‌سازی شبکه‌های اجتماعی

شبیه‌سازی، یکی از مهمترین و کارآمدترین رویکردهای تحلیل تعاملات در شبکه‌های اجتماعی است. شبیه‌سازی اجتماعی، برای آینده‌پژوهی، پشتیبانی تصمیم و مدیریت سیستم‌های اجتماعی در حوزه‌هایی مانند جامعه‌شناسی محاسباتی [۵]، اقتصاد و تجارت [۶، ۷، ۸]، نبرد، مسایل امنیتی و مدیریت بحران [۹، ۱۰]، بهداشت و

سلامت [۱۱]، سیاست [۱۲] و علوم رایانه [۱۳] مورد توجه بوده است. برخی از کاربردهای موفق شبیه‌سازی اجتماعی به عنوان نمونه در این بخش معرفی می‌شود.

گاهی برای بهبود و تنظیم روابط و نقش افراد در یک شبکه اجتماعی، از شبیه‌سازی استفاده می‌شود. در این کاربردها نقش افراد مختلف در شبکه اجتماعی تحلیل می‌شود و با شبیه‌سازی مدل‌های مختلف، الگوی بهتری برای اجرای فرایندها در شبکه اجتماعی ساخته می‌شود. مثلاً در [۱۳] مسأله جستجو و پرسش و پاسخ در یک شرکت خدمات مشاوره‌ای مدنظر قرار گرفته است. در این شرکت یک پرسش (درخواست مشاوره) به شرکت ارجاع می‌شود و با کمک شبکه اجتماعی کارمندان شرکت، فرد یا افراد مناسب برای پاسخگویی به سؤالات مشخص می‌شود. در این کاربرد روش‌ها و الگوریتم‌های مختلف برای پرسش و پاسخ، با کمک شبیه‌سازی در شبکه اجتماعی ارزیابی شده است. به بیان دیگر در این پژوهش روش‌های ممکن برای پیاده‌سازی نحوه ارجاع پرسش‌ها به پاسخ‌دهندگان شبیه‌سازی شده است و با کمک شبیه‌سازی، روش بهتر انتخاب شده است.

شبیه‌سازی برای تحلیل شبکه‌های اجتماعی مجازی نیز کاربردهای فراوانی دارد. به عنوان مثال فرض کنید بخواهیم برای تولید ابزارها و امکانات جدید در آینده برای یک شبکه اجتماعی مجازی (مانند فیس‌بوک) برنامه‌ریزی کنیم. برای تصمیم‌گیری در این شرایط، امکان شبیه‌سازی الگوی رشد شبکه اجتماعی، الگوی تعامل کاربران، الگوی نحوه انتشار، فراگیر شدن و محبوبیت امکانات و ابزارهای مختلف بسیار مفید خواهد بود. به عنوان مثال، در [۱۴] روشی برای شبیه‌سازی شبکه اجتماعی مجازی برخط ارائه شده است.

گاهی نتایج اتفاقات مختلف در یک محیط اجتماعی شبیه‌سازی می‌شود. اتفاقاتی مانند زلزله، آتش‌سوزی و تغییرات ترافیکی از این دسته هستند. مثلاً در [۹] حمله بیولوژیک به یک شهر در مدلی با نام BioWar در دانشگاه کارنگی ملون شبیه‌سازی شده است. این مدل به صورت یک سیستم چند عامله در یک شهر طراحی شده است و مواردی مانند واگیر بیماری، توزیع دارو و رفتارهای بهداشتی را شبیه‌سازی می‌کند.

گاهی از شبیه‌سازی برای بازنمایی صحنه جنگ استفاده می‌شود. در این کاربرد، ارتش خودی و ارتش دشمن به صورت شبکه‌های اجتماعی شبیه‌سازی می‌شوند و نتیجه اجرای تاکتیک‌ها و عملیات مختلف در نبرد بررسی می‌شود. به عنوان مثال در [۱۰] صحنه جنگ و نبرد بر اساس یک اتوماتای سلولی شبیه‌سازی شده است.

یکی از کاربردهای مهم شبیه‌سازی اجتماعی در مسایل مربوط به بازار است. مسایلی مانند اثربخشی تبلیغات، رفتار واکنشی مشتریان و مردم در بازار و نقش روابط بین مشتریان در رفتار آن‌ها در این حوزه جای می‌گیرند. مثلاً اگر بخواهیم کارآمدی یک روش تبلیغاتی را ارزیابی کنیم و یا یک روش تبلیغ مناسب طراحی کنیم، با کمک شبیه‌سازی می‌توانیم الگوی موردنظر تبلیغ را در یک شبکه شبیه‌سازی شده اجتماعی بیازماییم و با بررسی نتایج، الگوی مناسب را انتخاب کنیم. به عنوان مثال در [۶] رفتار مشتریان در یک بازار شبیه‌سازی شده است.

۴.۱ سیستم تصمیم‌یار مبتنی بر شبیه‌سازی اجتماعی

یک سیستم تصمیم‌یار، برنامه‌ای کامپیوتری است که از تصمیم‌گیری تجاری یا سازمانی پشتیبانی می‌کند. چنین سیستمی در زمان‌هایی که تصمیم‌گیری مشکل است، پارامترهای مختلفی در تصمیم‌گیری دخیل هستند و یا

فضای مسأله در حال تغییر است، بسیار مفید خواهد بود. در بسیاری از کاربردها، سیستم تصمیم‌یار با کمک یک پایگاه دانش و به خصوص یک انباره داده کار می‌کند و از مکعب‌های داده‌ای برای تحلیل داده‌ها و ساخت گزارش‌های آماری و تحلیلی استفاده می‌نماید [۱۵]. اما در این پژوهش از این ساختارها استفاده نمی‌شود و پشتیبانی از تصمیم با کمک شبیه‌سازی سناریوهای موردنظر مدیر در یک محیط چندعامله انجام می‌شود.

ارزش شبیه‌سازی شبکه‌های اجتماعی آن‌گاه که شبیه‌سازی در قالب یک سیستم تصمیم‌یار انجام شود، مشهودتر می‌گردد. در چنین سیستم تصمیم‌یاری، کاربر می‌تواند ویژگی‌های شبکه اجتماعی مورد نظر را تعیین کند و فرایندهای موجود در شبکه را مشخص کند. در نتیجه یک شبکه اجتماعی با مشخصات موردنظر شبیه‌سازی می‌شود و سناریوهای مختلف روی آن اجرا می‌گردد. به خصوص این که کاربر می‌تواند سناریوهای مختلف "چه می‌شود اگر" را اجرا و نتایج حاصل را بررسی کند. در پژوهش‌های بسیاری، بر نقش و اهمیت شبیه‌سازی اجتماعی در پشتیبانی تصمیم‌مدیران سیستم‌های اجتماعی تأکید شده است [۱۶، ۱۷، ۱۸، ۱۹، ۲۰].

سیستم تصمیم‌یاری که چشم‌انداز این پژوهش است، مجموعه‌ای از امکانات را برای شبیه‌سازی سیستم اجتماعی در اختیار مدیران قرار می‌دهد. ابتدا امکان توصیف شبکه هدف و ویژگی‌های این شبکه برای مدیران (کاربران) سیستم تصمیم‌یار فراهم می‌شود. سپس باید امکانی برای تولید شبکه مصنوعی با ویژگی‌های موردنظر فراهم شود. این شبکه مصنوعی، ممکن است با اندازه‌های بسیار بزرگ‌تر یا کوچک‌تر از شبکه هدف تولید شود. همچنین، امکانی برای ارزیابی شبکه مصنوعی ایجاد شده و محاسبه میزان شباهت آن با شبکه هدف لازم است. در مرحله بعدی، امکانی برای توصیف نحوه رفتار عامل‌ها در شبکه اجتماعی موردنیاز است. با کمک این امکان، مدیران رفتارهای مفروض برای اعضای شبکه را تعیین می‌کنند و نتیجه اجرای این رفتار را در کل شبکه شبیه‌سازی می‌کنند. در نهایت، امکانی برای اجرای نرم‌افزاری شبیه‌سازی به صورت کارآمد موردنیاز است. به‌ویژه، با پیدایش شبکه‌های بسیار بزرگ اجتماعی، امکان شبیه‌سازی مقیاس‌پذیر برای شبکه‌های بسیار بزرگ ضروری است.

هدف این رساله، حل برخی از مسایل پژوهشی برای ایجاد امکانات مذکور در یک سیستم تصمیم‌یار است. مسایل اصلی انتخاب شده در این پژوهش، در ادامه آمده است.

۵.۱ مسایل اساسی تحقیق

در این رساله دکتری، مسأله شبیه‌سازی فعالیت‌های اعضای یک شبکه اجتماعی مطرح شده است که برای تحلیل این فعالیت‌ها و به منظور کمک به تصمیم‌گیری بهتر مدیران این شبکه‌ها ساخته می‌شود. این صورت مسأله، یک مسأله یکپارچه است که دارای زیر مسایل پژوهشی مختلفی است. مسایل اصلی که در این پژوهش به آن‌ها پرداخته می‌شود، در ادامه آمده است.

- استخراج ویژگی‌های ساختاری مهم شبکه‌های اجتماعی. در این مسأله، از هر شبکه ویژگی‌ها و خواصی استخراج می‌شود که برای مقایسه، تحلیل و شبیه‌سازی شبکه‌ها مهم هستند. در جهت حل این مسأله،

توزیع درجه گره‌ها را به عنوان یکی از مهمترین ویژگی‌های ساختاری شبکه در نظر گرفتیم و روشی برای کمی‌سازی توزیع درجه در شبکه‌های اجتماعی ارائه کردیم.

- ارائه معیار شباهت ساختاری برای شبکه‌های اجتماعی. وجود معیاری برای مقایسه کمی ساختار شبکه‌ها برای شبیه‌سازی بسیار حایز اهمیت است. به‌ویژه، این معیار امکان مقایسه شبکه‌های مصنوعی تولید شده با شبکه‌های هدف و محاسبه میزان دقت شبکه‌های تولیدی را ارائه می‌کند. در این بخش، یک معیار شباهت برای شبکه‌های اجتماعی ارائه کردیم که دقیق‌تر از روش‌های موجود است.

- انتخاب مدل مولد مناسب برای تولید شبکه اجتماعی. متأسفانه هیچ مدل جامع و مانعی برای تولید شبکه اجتماعی وجود ندارد که در همه کاربردها مناسب باشد. به همین دلیل، در هر کاربرد، باید یک مدل مولد مناسب انتخاب کنیم که امکان تولید شبکه‌هایی مشابه با شبکه هدف را داشته باشد. در جهت حل این مسأله، دو روش هوشمند مختلف پیشنهاد کرده‌ایم.

- تخمین پارامترهای مناسب برای تنظیم مدل انتخاب شده. پس از انتخاب مدل مناسب، باید پارامترهای مناسب را نیز برای این مدل انتخاب کنیم تا شبکه‌هایی با ساختار موردنظر را تولید کند. برای حل این زیرمسأله نیز روشی دقیق و مقاوم نسبت به نویز ارائه کرده‌ایم.

- ارائه مدل و معماری نرم‌افزاری برای شبیه‌سازی شبکه‌های بزرگ. در نهایت، شبیه‌سازی باید توسط نرم‌افزاری کارا و مقیاس‌پذیر صورت پذیرد. چنین نرم‌افزاری باید دارای معماری مناسبی برای پشتیبانی از شبیه‌سازی‌های پیچیده، به‌ویژه برای شبکه‌های بسیار بزرگ باشد. در این مسأله نیز به ارائه یک مدل مرجع اولیه نرم‌افزاری برای شبیه‌سازی توزیع‌شده پرداخته‌ایم.

مسائل مطرح شده، مسایلی مستقل نیستند و دارای یک هدف و دورنمای مشترک (تولید سیستم تصمیم‌یار) هستند. همچنین، راه‌حل هر یک از مسائل این پژوهش، به عنوان سنگ‌بنای مسأله بعدی در همین پژوهش مورد استفاده قرار می‌گیرد.

۶.۱ ساختار رساله

ساختار این رساله در ادامه به این شرح است: فصل ۲ به مرور ادبیات موضوع و کارهای مرتبط می‌پردازد. فصل ۳ به توصیف مجموعه شبکه‌های واقعی و مصنوعی می‌پردازد که در این رساله از آن‌ها استفاده شده است. در فصل ۴ روشی برای استخراج ویژگی از شبکه‌های اجتماعی ارائه می‌شود. در فصل ۵ روشی برای مقایسه ویژگی‌های ساختاری شبکه‌ها ارائه می‌شود. در فصل ۶ روشی برای انتخاب مدل و تنظیم پارامترها جهت تولید شبکه اجتماعی مصنوعی پیشنهاد می‌شود و اهمیت و دقت آن ارزیابی و تبیین می‌گردد. نهایتاً در فصل ۷ این رساله جمع‌بندی می‌شود و چشم‌انداز کارهای آینده این پژوهش بیان می‌شود. همچنین پیوست الف، یک مدل مرجع اولیه نرم‌افزاری برای شبیه‌سازی مقیاس‌پذیر برای شبکه‌های اجتماعی پیشنهاد می‌کند.

۲ مرور ادبیات موضوع

۱.۲ ویژگی‌های ساختاری معمول در شبکه‌های اجتماعی

بسیاری از شبکه‌های اجتماعی دارای ویژگی‌های ساختاری مشترکی هستند. هر چند شبکه‌های اجتماعی در حوزه‌ها و کاربردهای مختلفی ایجاد می‌شوند، اما این ویژگی‌های مشترک در بیشتر این شبکه‌ها مشاهده می‌شوند. مثلاً برخی ویژگی‌ها که در گراف ساختار صفحات وب دیده می‌شود، در گراف ساختار ارتباطات افراد در شبکه تلفن همراه هم دیده می‌شود. مطالعه ویژگی‌های مشترک شبکه‌های اجتماعی، بخش مهمی از این پژوهش بوده است. به خصوص برای تولید شبکه اجتماعی مصنوعی، باید ویژگی‌های مهم ساختار شبکه‌های اجتماعی لحاظ شود. برخی از مهمترین ویژگی‌های مشترک این شبکه‌ها در ادامه آمده است. لازم به ذکر است که در این پژوهش، فقط گراف‌های ساده (غیروزدار و غیرجهت‌دار) در نظر گرفته شده‌اند.

۱.۱.۲ گراف تُنک

شبکه‌های اجتماعی، معمولاً گراف‌هایی تُنک^۱ هستند [۲۱] و در آن‌ها نسبت بسیار کمی از یال‌های ممکن، ایجاد شده‌اند. گراف تُنک در مقابل گراف چگال^۲ قرار می‌گیرد که در آن، یال‌های ایجاد شده به حداکثر یال‌های ممکن نزدیک هستند. برخی از معیارهایی که برای اندازه‌گیری این خاصیت در شبکه‌ها محاسبه می‌شوند عبارتند از متوسط درجه رئوس^۳ و چگالی^۴ شبکه.

۲.۱.۲ دنیای کوچک

دنیا در شبکه‌های اجتماعی کوچک است. اگر شبکه را به صورت یک گراف بازنمایی کنیم، فاصله گره‌ها در این گراف بسیار کم است و حتی در شبکه‌های بسیار بزرگ اجتماعی، میانگین فاصله بین گره‌ها حدود شش یال است. این ویژگی که از آن تحت عنوان "شش درجه جدایی" یاد می‌شود، بدین معناست که در گراف ارتباطات بین افراد، هر دو فرد دلخواه در دنیا به طور میانگین با هم شش "دوست" فاصله دارند. اولین بار یک روانشناس اجتماعی با نام میلگرم در سال ۱۹۶۷ این ویژگی را مورد مطالعه قرار داد [۲۲]. وی در دوره‌ای که امکاناتی

^۱ Sparse

^۲ Dense

^۳ Average Degree

^۴ Density

مانند ایمیل و تلفن همراه وجود نداشت، افرادی را از ایالت‌های مختلف آمریکا به صورت تصادفی انتخاب کرد و از آن‌ها خواست تا یک نامه را به دست فردی دیگر برسانند و اگر وی را به صورت مستقیم نمی‌شناسند، نامه را به کسی بدهند که احتمال دارد او را بشناسد. در این مطالعه، نامه‌هایی که به مقصد رسیدند به طور متوسط با فاصله شش نفر از نفر مبدأ به نفر مورد نظر رسیدند. هر چند تعداد کمی از نامه‌ها در این آزمایش به مقصد رسیدند (فقط ۶۴ نامه از مجموع ۲۹۶ نامه) اما آزمایش‌های بعدی و جدیدتر نه تنها بر ویژگی دنیای کوچک صحنه می‌گذارند، بلکه حتی عدد شش را به عنوان حدود فاصله بین افراد در گراف‌های اجتماعی تأیید می‌کنند [۲۳، ۲۴]. فاصله بین دو گره در یک گراف با اندازه کوتاه‌ترین مسیر بین این دو گره تعریف می‌شود. ویژگی "دنیای کوچک" به معنای کوچک بودن اندازه میانگین فاصله گره‌ها در شبکه‌های اجتماعی است. معمولاً گرافی شامل این ویژگی در نظر گرفته می‌شود که میانگین فاصله گره‌ها در آن با سرعتی بیشتر از لگاریتم تعداد گره‌ها رشد نکند [۲۵] (رابطه ۱۰۲ را ببینید). اهمیت این ویژگی زمانی بیشتر می‌شود که دقت کنیم شبکه‌های اجتماعی معمولاً گراف‌هایی تُنک هستند و چگالی پایینی دارند.

$$l = \log(N); \quad N \rightarrow \infty \quad (1.2)$$

برخی از معیارهایی که برای اندازه‌گیری خاصیت دنیای کوچک در شبکه‌ها محاسبه می‌شوند عبارتند از میانگین فاصله گره‌ها^۵، قطر شبکه^۶، و قطر مؤثر^۷.

۳.۱.۲ توزیع درجه گره‌ها

در بسیاری از شبکه‌های اجتماعی واقعی، توزیع درجه رئوس گراف، دارای ظاهری "دُم دراز"^۸ است. به خصوص توزیع توانی^۹ برای بسیاری از شبکه‌ها گزارش شده است [۲۶]. در توزیع توانی تعداد گره‌هایی که درجه d دارند (N_d) از رابطه $N_d \propto d^{-\gamma}$ پیروی می‌کند که در آن γ یک عدد مثبت است. شبکه‌هایی با این ویژگی را شبکه‌های "مستقل از مقیاس"^{۱۰} می‌گویند [۲]. نمودار توزیع توانی در شکل ۱۰۲ آمده است. البته برای نمایش توزیع توانی، معمولاً نمودار توزیع درجه گره‌ها را به صورت "نمودار لاگ-لاگ" رسم می‌کنند که در آن $\log(N_d)$ براساس $\log(d)$ رسم می‌شود. همان‌طور که رابطه ۲.۲ نشان می‌دهد، این نمودار برای توزیع توانی به صورت خطی با شیب $-\gamma$ خواهد بود. شکل ۲.۲ نمودار توزیع درجه رئوس را برای گراف صفحات وب نشان می‌دهد [۱]. به این حالت از توزیع درجه گره‌ها، حالت دُم‌دراز گفته می‌شود زیرا در نمودار N_d براساس d ، انتهای نمودار بسیار طولانی است و برای درجات بالای گره‌ها همچنان تعدادی گره (تعداد کم) دیده می‌شود [۲۷].

^۵Average shortest path length

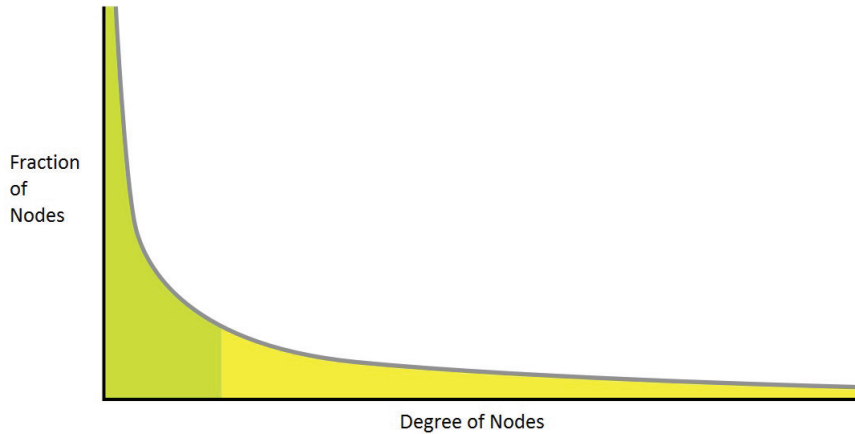
^۶Diameter

^۷Effective diameter

^۸Long-tail degree distribution

^۹Power-law degree distribution

^{۱۰}Scale free



شکل ۱.۲: توزیع درجه نمایی در شبکه‌های مستقل از مقیاس

$$N_d \propto d^{-\gamma}$$

$$N_d = c.d^{-\gamma} \quad (۲.۲)$$

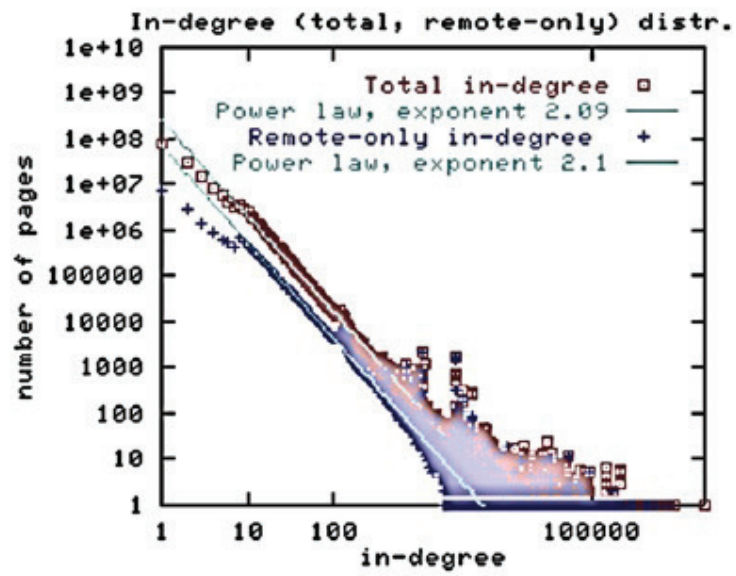
$$\log(N_d) = \log(c) - \gamma \log(d)$$

در بسیاری از روش‌های کاوش و تحلیل شبکه‌ها، نیازمند وجود تعداد محدودی عدد به عنوان نماینده و توصیف‌گر توزیع درجه هستیم. برای این منظور گاهی توزیع درجه شبکه مدنظر را بر روی یک توزیع توانی مدل می‌کنند و از γ (نمای توزیع توانی) به عنوان نماینده توزیع درجه استفاده می‌کنند. در فصل ۴ به بررسی روش‌های موجود در این زمینه می‌پردازیم و یک روش جدید نیز ارائه می‌کنیم.

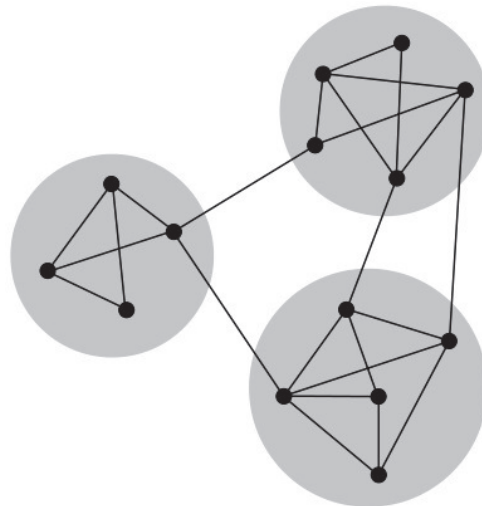
۴.۱.۲ ساختار گروهی شبکه

بسیاری از شبکه‌های اجتماعی ساختاری پیمانانه‌ای و گروهی دارند. در این شبکه‌ها گروه‌ها و خوشه‌هایی مشاهده می‌شود که ارتباط اعضا درون این خوشه‌ها بیشتر از ارتباط آن‌ها با بیرون خوشه‌ها است. در شبکه‌های اجتماعی مفاهیمی مثل گروه دوستی، همکاران، اقوام و غیره به شکل‌گیری خوشه‌ها و گروه‌ها منجر می‌شوند. راه‌حل ساده‌ای برای اندازه‌گیری میزان تطابق یک شبکه با ساختار گروهی وجود ندارد. یک راه حل برای این کار خوشه‌بندی شبکه با استفاده از الگوریتم‌های خوشه‌بندی گراف و سپس محاسبه میزان اصالت خوشه‌های محاسبه شده است. اصالت خوشه‌ها زمانی بالا در نظر گرفته می‌شود که ارتباط گره‌ها در درون خوشه‌ها بیشتر و چگال‌تر از ارتباط گره‌ها بین خوشه‌ها باشد. برای خوشه‌بندی گراف الگوریتم‌های مختلفی وجود دارد و یکی از بهترین و پر استفاده‌ترین معیارها برای محاسبه اصالت خوشه‌ها نیز معیار پیمانگی^{۱۱} است [۲۸]. شکل ۳.۲ نمایی شهودی از یک شبکه با ساختار گروهی را نشان می‌دهد.

^{۱۱}Modularity



شکل ۲.۲: توزیع توانی درجه رهوس در گراف صفحات وب [۱]



شکل ۳.۲: شبکه‌ای با ساختار گروهی

۵.۱.۲ ویژگی تراگذری یال‌ها

در شبکه‌های واقعی و به خصوص شبکه‌های اجتماعی، گره‌ها تمایل به شرکت در خوشه‌ها دارند و معمولاً در گراف ارتباطات، مثلث‌هایی از گره‌های متصل می‌سازند [۲۷، ۲]. از نظر شهودی، وقتی فرد "الف" با "ب" ارتباط دارد (مثلاً با هم دوست هستند) و "ب" هم با "ج" ارتباط دارد، احتمال برقراری ارتباط بین "الف" و "ج" بالا خواهد بود و احتمالاً به مرور زمان در گراف ارتباطات، روابطی مانند \wedge به شکل \triangle درمی‌آیند. معیارهایی مانند "ضریب خوشه‌بندی"^{۱۲} و تراگذری^{۱۳}، که ناظر بر این ویژگی در شبکه‌ها هستند، در شبکه‌های اجتماعی بیشتر از شبکه‌های کاملاً تصادفی هستند.

۶.۱.۲ هم‌رنگی

یکی از ویژگی‌های شبکه‌های اجتماعی وجود شباهت بین گره‌های متصل به هم است. به عنوان مثال در یک شبکه دوستی، دو دوست احتمالاً تشابهات و نقاط مشترک زیادی دارند. به این ویژگی، هم‌رنگی^{۱۴} گفته می‌شود (کبوتر با کبوتر باز با باز، کند هم‌جنس با هم‌جنس پرواز^{۱۵}). یکی از جلوه‌های هم‌رنگی در شبکه‌های اجتماعی، جورش درجه‌ها^{۱۶} است که میزان تناسب درجه گره‌های متصل به هم را نشان می‌دهد. به عبارت دیگر، این ویژگی همبستگی بین درجه گره‌هایی را نشان می‌دهد که با یال به هم متصل شده‌اند. از منظر هم‌رنگی، این ویژگی میزان شباهت گره‌های متصل به هم را از نظر درجه گره‌ها نشان می‌دهد. معیاری با نام "جورش درجه‌ها" برای محاسبه این ویژگی وجود دارد.

۲.۲ مدل‌های تولید شبکه اجتماعی تصادفی

در کنار تشخیص ویژگی‌ها و قوانینی که بر ساختار شبکه‌های اجتماعی حاکم هستند، روش‌ها و مدل‌هایی برای تولید شبکه‌های اجتماعی توسعه یافته‌اند. این مدل‌ها شبکه‌های اجتماعی تصادفی را به نحوی می‌سازند که شبکه مصنوعی ایجاد شده دارای ویژگی‌های عمومی شبکه‌های اجتماعی واقعی باشند. ایجاد روش‌هایی برای تولید شبکه تصادفی به خصوص از این نظر مهم است که امکان شبیه‌سازی، تحلیل "چه می‌شود اگر" و بررسی فرضیه‌ها را در شبکه‌هایی که امکان جمع‌آوری داده‌ها از آن‌ها سخت یا غیرممکن است، فراهم می‌سازند [۲۹]. مثلاً برای پاسخ به این سؤال که "انتشار یک خبر پنج سال دیگر در شبکه تلفن همراه کشور با چه سرعتی انجام می‌شود؟" ساخت شبکه تصادفی با ویژگی‌های شبکه تلفن همراه (اما چند برابر بزرگتر از شبکه امروزی) و شبیه‌سازی این شبکه مفید خواهد بود. روش‌های مختلفی برای تولید شبکه تصادفی پیشنهاد شده است. هر یک از این روش‌ها، که غالباً از مدل‌های ریاضی و احتمالاتی و یا روش‌های مبتنی بر عامل برای تولید گراف

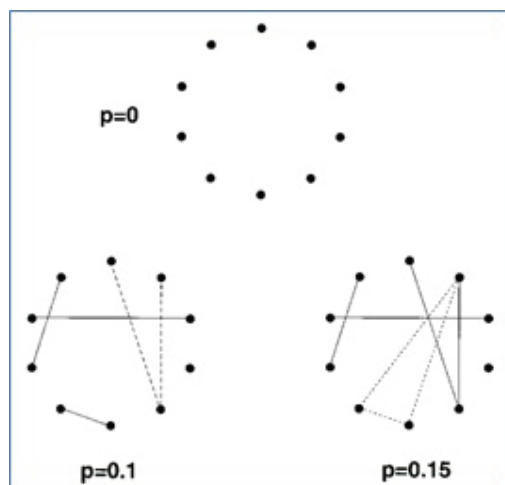
^{۱۲}Clustering Coefficient

^{۱۳}Transitivity

^{۱۴}Homophily

^{۱۵}Birds of a feather flock together

^{۱۶}Assortativity



شکل ۴.۲: نمونه گراف‌های تصادفی که با مدل ER ساخته شده‌اند [۲]

استفاده می‌کنند، "مدل مولد شبکه" ^{۱۷}، یا به سادگی "مدل شبکه" خوانده می‌شود. در این بخش، به برخی از مدل‌های مهم و پرکاربرد در این زمینه اشاره می‌کنیم.

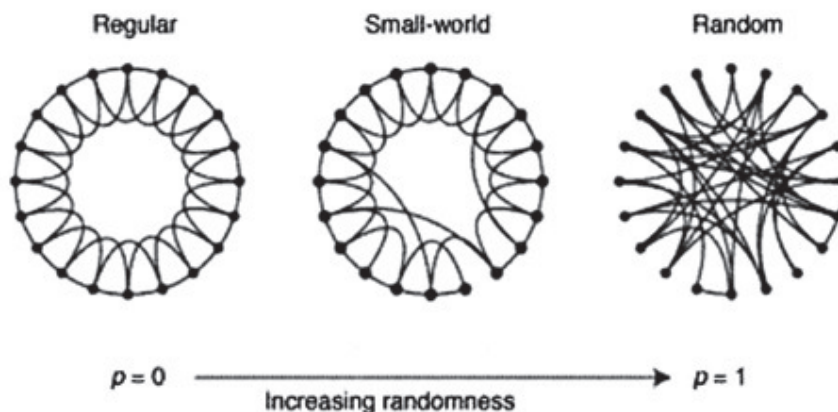
ساده‌ترین مدل مولد گراف، مدل Erdos-Rényi (ER) است [۳۰]. در این مدل بین هر دو گره دلخواه، با احتمالی یکسان و مستقل یک یال اضافه می‌شود. شکل ۴.۲ نمونه‌هایی از گراف‌های تولید شده توسط این مدل را نشان می‌دهد. هر چند این مدل از نظر ریاضی بسیار مورد توجه بوده است، اما گراف‌های تولید شده از روی این مدل ویژگی‌های شبکه‌های واقعی را ندارند. به عنوان مثال توزیع درجه گره‌ها در این مدل از توزیع دو جمله‌ای تبعیت می‌کند که شبیه توزیع توانی (ویژگی دم‌دراز) نیست. همچنین ویژگی جورش درجه‌ها ^{۱۸} در این گراف‌ها دیده نمی‌شود. بهبودها و تغییرات مختلفی در مدل ER ایجاد شده است. یکی از این تغییرات، که برخلاف مدل ER توزیع درجه توانی ایجاد می‌کند، به نام مدل Random-powerlaw شناخته می‌شود [۳۱]. در این مدل توان توزیع درجه توانی (γ) قابل تنظیم است.

مدل کلاسیک دیگری که برای تولید شبکه تصادفی ارائه شده است، مدل Barabási-Albert (BA) است [۳۲]. هدف این مدل تولید گرافی است که ویژگی توزیع درجه توانی را داشته باشد. در این مدل هر گره جدید تمایل به برقراری ارتباط با گره‌هایی دارد که درجه بیشتری داشته باشند. به این مدل گسترش گراف، "اتصال ترجیحی" ^{۱۹} گفته می‌شود. در این مدل، در هر مرحله، یک گره به گراف اضافه می‌شوند و احتمال برقراری ارتباط بین گره جدید و یک گره قدیمی، متناسب با درجه گره قدیمی است. بنابراین به مرور زمان گره‌هایی که درجه بالایی دارند، درجه بالاتری پیدا می‌کنند. از این ویژگی تحت عنوان خاصیت "ثروتمند ثروتمندتر می‌شود"

^{۱۷}Network Generative Model

^{۱۸}Assortativity

^{۱۹}Preferential Attachment



شکل ۵.۲: گراف‌های ایجاد شده توسط مدل دنیای کوچک Watts-Strogatz [۲]

^{۲۰} برای این مدل یاد می‌شود. تعداد یال‌هایی که در هر مرحله به گره جدید اضافه می‌شود قابل تنظیم است. یکی دیگر از مدل‌های تولید گراف تصادفی، مدل دنیای کوچک ^{۲۱} Watts-Strogatz است [۲۵]. هدف اصلی این مدل تولید گراف‌هایی با ویژگی‌های “دنیای کوچک” و “تراگذری یال‌ها” است. در این مدل برای تولید یک گراف ابتدا یک حلقه از N گره در نظر گرفته می‌شود که در آن هر گره به k همسایه اطراف خود از هر جهت (مجموعاً به $2k$ گره) متصل است. سپس هر یال در این گراف با احتمال p تغییر می‌کند و یک سر آن به یک گره دیگر – که به احتمال یکنواخت از بین سایر گره‌ها انتخاب می‌شود – متصل می‌شود. در این فرایند از ایجاد حلقه بر روی گره‌ها خودداری می‌شود. اگر مقدار p به صفر نزدیک باشد، تراگذری یال‌ها و ضریب خوشه‌بندی مقادیری بالا خواهند داشت، اما فاصله بین گره‌ها هم به طور متوسط بالا خواهد بود. در مقابل اگر مقدار p به یک نزدیک شود، گراف به یک گراف تصادفی با قطر کم اما ضریب خوشه‌بندی پایین تبدیل می‌شود. واتس و استروگاتس نشان دادند که به ازای مقادیر میانی برای p هر دو ویژگی در گراف نهایی حفظ می‌شود. این وضعیت در شکل ۵.۲ نشان داده شده است. در این شکل، گراف‌های ایجاد شده توسط مدل دنیای کوچک به ازای $k=2$ و $N=20$ و مقادیر مختلف p نشان داده شده‌اند.

یکی دیگر از مدل‌های تولید شبکه‌های اجتماعی، استفاده از یک عملگر ماتریسی با نام “حاصلضرب کرونیکر” است [۲۹]. با کمک این عملگر ماتریس مجاورت گراف به صورت بازگشتی و خودمشابه ساخته می‌شود. اگر A یک ماتریس $m \times n$ و B یک ماتریس $p \times q$ باشد حاصلضرب کرونیکر این دو ماتریس به صورت یک ماتریس $mp \times nq$ خواهد بود. در این رویکرد، برای تولید یک شبکه تصادفی بزرگ یک گراف اولیه (K_1) را بارها در خودش ضرب می‌کنیم و به یک ماتریس مجاورت برای گرافی بزرگ می‌رسیم. اگر گراف اولیه برخی شرایط لازم را داشته باشد (مانند متصل بودن و داشتن حلقه روی همه گره‌ها) گراف نهایی که از این روش ساخته می‌شود بسیاری از ویژگی‌های عمومی شبکه‌های اجتماعی حقیقی را خواهد داشت (مانند دنیای

^{۲۰} Rich get richer

^{۲۱} Small World

کوچک، توزیع درجه دُم‌دراز، چگال شدن شبکه و کاهش اندازه قطر). شبکه‌ای که در این مدل ساخته می‌شود، به خاطر روش تولید آن، حالتی خودمشابه و فرکتالی دارد. در مدل بهبودیافته حاصلضرب کرونیکر [۲۹]، به جای این که ماتریس مجاورت باینری باشد و از صفر و یک تشکیل شود، از اعداد حقیقی بین صفر و یک تشکیل می‌شود. به این مدل "ماتریس کرونیکر تصادفی" گفته می‌شود. برای ساخت شبکه از روی یک ماتریس کرونیکر تصادفی، اگر در خانه $[u, v]$ از این ماتریس عدد p قرار گرفته باشد، در گراف متناظر به احتمال p بین u و v یال رسم می‌شود. به این عمل تحقق گراف ^{۲۲} گفته می‌شود.

مدل آتش در جنگل ^{۲۳} یکی از مدل‌های مهم تولید شبکه اجتماعی است که در آن یال‌ها با فرایندی مشابه با انتشار آتش گسترش می‌یابند [۳۳]. احتمال آتش‌سوزی روبه‌جلو (p) و احتمال آتش‌سوزی روبه‌عقب (p_b) در این مدل قابل تنظیم هستند. در این مدل، در هر مرحله یک گره به گراف اضافه می‌شود. ابتدا، گره جدید (v) یکی از گره‌های موجود (w) را به صورت تصادفی انتخاب می‌کند و به آن متصل می‌شود. سپس تعدادی از همسایگان w هم انتخاب می‌شوند و به v متصل می‌شوند. این مدل ویژگی‌هایی مانند کاهش اندازه قطر ^{۲۴} و چگال شدن شبکه ^{۲۵} را پشتیبانی می‌کند.

مدل‌های شبکه به موارد ذکر شده محدود نیست و تاکنون مدل‌های مختلف و متنوعی برای تولید شبکه‌های اجتماعی پیشنهاد شده است [۳۰، ۳۲، ۳۳، ۲۹، ۳۱، ۳۴، ۳۵، ۳۶، ۳۷، ۳۸، ۳۹، ۴۰، ۴۱، ۴۲، ۴۳، ۴۴]. هرچند برخی از مدل‌ها دایره وسیع‌تری از ویژگی‌های شبکه‌های حقیقی را پشتیبانی می‌کنند، نمی‌توانیم یک روش را به صورت مطلق به عنوان "بهترین" انتخاب کنیم. در واقع هر یک از مدل‌های موجود در برخی کاربردها مناسب و در کاربردهای دیگر نامناسب هستند. یکی از چالش‌های موجود برای تولید شبکه‌های اجتماعی، انتخاب یک مدل مناسب برای کاربرد موردنظر است. چالش دیگر، بعد از انتخاب مدل، انتخاب پارامترهای مناسب برای تولید شبکه موردانتظار است. در واقع هر مدل، با کمک پارامترهایی تنظیم می‌شود و در هر کاربرد، با توجه به ساختار موردنظر برای شبکه مصنوعی، باید پارامترهای مدل را تنظیم کنیم. درباره این دو چالش در فصل ۶ بحث می‌کنیم و روش‌های جدیدی ارائه می‌کنیم.

^{۲۲}Realization^{۲۳}Forest-Fire (FF)^{۲۴}Shrinking Diameter^{۲۵}Densification

۳ مجموعه داده‌ها

در این فصل به معرفی مجموعه داده‌های^۱ استفاده شده در این پژوهش می‌پردازیم. این مجموعه داده‌ها شامل شبکه‌های حقیقی^۲، شبکه‌های مصنوعی^۳ و شبکه‌های زمان‌دار^۴ است. این مجموعه داده‌ها برای ارزیابی روش‌های ارائه شده در این پژوهش مورد استفاده قرار گرفته‌اند. هرچند پژوهش انجام شده برای شبکه‌های وزن‌دار و جهت‌دار نیز قابل‌گسترش است، در این پژوهش فقط گراف‌های ساده (بدون وزن و بدون جهت) در نظر گرفته شده‌اند و مجموعه داده آماده شده فقط شامل گراف‌های ساده است.

۱.۳ شبکه‌های حقیقی

به عنوان شبکه‌های حقیقی، مجموعه‌ای از ۳۳ شبکه از دنیای واقعی در انواع مختلف جمع‌آوری شده است. این شبکه‌ها از شش نوع مختلف انتخاب شده است: شبکه دوستی^۵، شبکه‌های ارتباطی^۶، شبکه‌های همکاری^۷، شبکه‌های ارجاعات^۸، شبکه‌های نظیر به نظیر^۹ و شبکه‌های صفحات وب^{۱۰}. نوع شبکه نشانه‌ای از شباهت بین شبکه‌های هم‌نوع است: شبکه‌های هم‌نوع معمولاً با فرایندهای مشابهی شکل می‌گیرند و ساختارهای مشابهی تولید می‌کنند.

جدول ۱.۳ مجموعه داده شبکه‌های حقیقی را نشان می‌دهد. بیشتر این شبکه‌ها در دسترس عموم قرار دارند و در پژوهش‌های مختلفی، قبلاً استفاده شده‌اند. در این میان، دو شبکه `CiteSeerX` و `Col-lab_CiteSeerX` را با کمک یک خزنده^{۱۱} از کتابخانه دیجیتال `CiteSeerX` [۴۵] استخراج کرده‌ایم. این دو شبکه به عنوان دو مجموعه داده مجزا نیز (مجموعه داده‌های زمان‌دار) در ادامه معرفی خواهند شد.

^۱dataset

^۲Real Networks

^۳Artificial Networks

^۴Temporal Networks

^۵Friendship Networks

^۶Communication Networks

^۷Collaboration Networks

^۸Citation Networks

^۹Peer-to-Peer Networks

^{۱۰}Web Graphs

^{۱۱}Crawler

جدول ۱.۳: مجموعه شبکه‌های حقیقی

مرجع	تعداد یال	تعداد گره	شناسه شبکه	نوع شبکه
[۴۶] SNAP	۴۲۰,۸۹۹	۳۴,۵۴۶	Cit-HepPh	شبکه‌های ارجاعات
[۴۶] SNAP	۳۵۲,۳۰۴	۲۷,۷۷۰	Cit-HepTh	
[۴۷] DBLP	۲,۲۸۴,۶۹۴	۴۷۵,۸۸۶	dblp cite	
[۴۵] CiteSeerX	۱۱,۷۹۱,۲۲۸	۱,۱۰۶,۴۳۱	Cit CiteSeerX	
[۴۶] SNAP	۱۹۸,۰۸۰	۱۸,۷۷۲	CA-AstroPh	شبکه‌های همکاری
[۴۶] SNAP	۹۳,۴۶۵	۲۳,۱۳۳	CA-CondMat	
[۴۶] SNAP	۲۵,۹۸۵	۹,۸۷۷	CA-HepTh	
[۴۵] CiteSeerX	۵,۳۱۳,۱۰۱	۱,۲۶۰,۲۹۲	Collab CiteSeerX	
[۴۶] SNAP	۱,۰۴۹,۸۶۶	۳۱۷,۰۸۰	com-dblp	
[۴۷] DBLP	۳,۴۸۹,۵۷۲	۹۷۵,۰۴۴	dblp collab	
[۴۸] Sommer	۱,۸۷۱,۰۷۰	۵۱۱,۱۶۳	dblp20080824	
[۴۹] Rossetti	۱۶,۶۷۹	۴,۱۵۵	IMDB-09	
[۴۶] SNAP	۱۴,۴۹۰	۵,۲۴۲	CA-GrQc	
[۴۶] SNAP	۱۱۸,۵۰۵	۱۲,۰۰۸	CA-HepPh	
[۵۰] Arenas	۵,۴۵۱	۱,۱۳۳	EmailURV	شبکه‌های ارتباطات
[۵۱, ۴۶] SNAP	۱۸۳,۸۳۱	۳۶,۶۹۲	Email-Enron	
[۵۲] Konect	۳۶۵,۰۲۵	۲۶۵,۲۱۴	Email-EuAll	
[۴۶] SNAP	۴,۶۵۹,۵۶۵	۲,۳۹۴,۳۸۵	WikiTalk	
[۵۳] NetData	۱۵۹	۶۲	Dolphins	شبکه‌های دوستی
[۵۴] MaxPlanck	۸۱۷,۰۹۰	۶۳,۷۳۱	facebook-links	
[۴۶] SNAP	۵۰۷,۸۳۳	۷۷,۳۶۰	Slashdot0811	
[۴۶] SNAP	۵۴۳,۳۸۱	۸۲,۱۶۸	Slashdot0902	
[۴۶] SNAP	۴۰۵,۷۴۰	۷۵,۸۷۹	soc-Epinions1	
[۴۹] Rossetti	۸,۵۹۳	۲,۵۶۶	Twitter-Richmond	
[۵۴] MaxPlanck	۲,۹۹۰,۴۴۳	۱,۱۳۸,۴۹۹	youtube-d-growth	
[۴۶] SNAP	۶,۶۴۹,۴۷۰	۶۸۵,۲۳۰	web-BerkStan	شبکه‌های صفحات وب
[۴۶] SNAP	۴,۳۲۲,۰۵۱	۸۷۵,۷۱۳	web-Google	
[۴۶] SNAP	۱,۱۰۳,۸۳۵	۳۲۵,۷۲۹	web-NotreDame	
[۴۶] SNAP	۱,۹۹۲,۶۳۶	۲۸۱,۹۰۳	web-Stanford	
[۴۶] SNAP	۳۹,۹۹۴	۱۰,۸۷۶	p2p-Gnutella04	شبکه‌های نظیر به نظیر
[۴۶] SNAP	۳۱,۸۳۹	۸,۸۴۶	p2p-Gnutella05	
[۴۶] SNAP	۳۱,۵۲۵	۸,۷۱۷	p2p-Gnutella06	
[۴۶] SNAP	۲۰,۷۷۷	۶,۳۰۱	p2p-Gnutella08	

۲.۳ شبکه‌های مصنوعی

شبکه‌های مصنوعی شبکه‌هایی هستند که در یک کاربرد حقیقی به وجود نیامده‌اند، بلکه توسط مدل‌های مولد ایجاد شده‌اند. در این بخش، ۶۰۰۰ شبکه مصنوعی با استفاده از شش مدل مختلف تولید شدند. مدل‌های مورد استفاده روش‌های پرکاربردی برای تولید شبکه‌های مصنوعی هستند. این مدل‌ها برای تولید گراف‌هایی با ساختارهای مشابه با شبکه‌های واقعی استفاده می‌شوند. شش مدل موجود در مجموعه داده شبکه‌های مصنوعی عبارتند از: مدل Barabási-Albert [۳۲]، مدل Erdos-Rényi [۳۰]، مدل Forest-Fire [۳۳]، مدل Kronecker [۲۹]، مدل Random-powerlaw [۳۱] و مدل Small-world (Watts-Strogatz) [۲۵]. مجموعه شبکه‌های ایجاد شده اندازه‌های متنوعی دارند. تعداد گره‌ها در هر یک از این شبکه‌ها بین ۱۰۰۰ تا ۵۰۰۰ گره با متوسط ۲،۹۳۶ گره در هر نمونه است. متوسط تعداد یال‌ها نیز ۱۳،۷۱۴ است. در این مجموعه داده، مدل هر شبکه شاهدهی مبنی بر شباهت آن شبکه با شبکه‌های هم‌مدل است، زیرا شبکه‌هایی که از یک مدل تولید شده‌اند، فرایند مشابهی را برای تولید طی کرده‌اند و ساختاری مشابه دارند.

مدل‌های استفاده‌شده در این مجموعه داده، مدل‌هایی مهم و پر استفاده برای تولید شبکه‌های مصنوعی هستند که طیف گسترده‌ای از ساختارهای شبکه‌ای را پوشش می‌دهند. این مدل‌ها همراه با پارامترهای استفاده شده برای تولید شبکه‌های مصنوعی در ادامه توضیح داده شده‌اند. مقادیر مورد استفاده برای پارامترهای این مدل‌ها، با توجه به نکات و توصیه‌های مطرح شده در مراجع اصلی آن‌ها انتخاب شده‌اند. از هر یک از این مدل‌ها، هزار شبکه مصنوعی تولید شده است و در مجموعه داده شبکه‌های مصنوعی قرار گرفته است.

- **مدل Barabási-Albert (BA)**. این مدل کلاسیک، شبکه‌هایی مستقل از مقیاس^{۱۲} ایجاد می‌کند که بر مبنای فرایند اتصال ترجیحی^{۱۳} ایجاد شده‌اند و توزیع درجه توانی دارند [۳۲]. در این مدل، گره‌ها به صورت تدریجی (یکی پس از دیگری) به گراف اضافه می‌شوند. هر گره جدید به صورت تصادفی به m گره موجود در شبکه متصل می‌شود و احتمال انتخاب هر گره موجود برای اتصال متناسب با درجه آن است (اتصال ترجیحی). در مجموعه شبکه‌های مصنوعی، مقدار m به صورت تصادفی بین یک تا ده انتخاب شده است ($1 \leq m \leq 10$).

- **مدل Erdos-Rényi (ER)**. این مدل شبکه‌های کاملاً تصادفی تولید می‌کند [۳۰]. چگالی شبکه در این مدل قابل تنظیم است. چگالی هر شبکه با نسبت تعداد یال‌های موجود به یال‌های ممکن تعریف می‌شود. در مجموعه داده شبکه‌های مصنوعی، چگالی شبکه‌های ER در محدوده $0.002 \leq density \leq 0.005$ قرار دارند.

- **مدل Forest-Fire (FF)**. در این مدل، یال‌ها با فرایندی مشابه با انتشار آتش گسترش می‌یابند [۳۳].

^{۱۲}Scale Free

^{۱۳}Preferential Attachment

این مدل ویژگی‌هایی مانند کاهش اندازه قطر^{۱۴} و چگال شدن شبکه^{۱۵} را پشتیبانی می‌کند. این مدل توسط دو پارامتر تنظیم می‌شود: احتمال آتش‌سوزی روبه‌جلو (p) و احتمال آتش‌سوزی روبه‌عقب (p_b). برای تولید شبکه‌های مصنوعی، احتمال p_b را ثابت در نظر گرفتیم ($p_b = 0.32$) و مقدار p را به صورت تصادفی از بازه $0 \leq p \leq 0.3$ انتخاب کردیم.

● **مدل گراف‌های (KG) Kronecker**. این مدل با استفاده از یک عملگر ماتریسی (ضرب کرونیکر) روی یک ماتریس اولیه کوچک شبکه نهایی را به تدریج می‌سازد [۲۹]. این مدل، ویژگی‌هایی مانند دنیای کوچک، توزیع درجه دُم‌دراز، چگال شدن شبکه و کاهش اندازه قطر را پشتیبانی می‌کند. همچنین برخی خواص این مدل به صورت ریاضی قابل اثبات است. گراف‌هایی که از این مدل در مجموعه داده شبکه‌های مصنوعی گنجانده شده‌اند، بر اساس یک گراف اولیه 2×2 ایجاد شده‌اند. چهار درایه این ماتریس به صورت تصادفی انتخاب شده‌اند. بازه انتخاب تصادفی این درایه‌ها عبارتند از: $0.7 \leq P_{1,1} \leq 0.9$ ، $0.5 \leq P_{1,2} \leq 0.7$ ، $0.4 \leq P_{2,1} \leq 0.6$ ، $0.4 \leq P_{2,2} \leq 0.4$ ، (درایه $P_{i,j}$ به معنی سطر i و ستون j از ماتریس است).

● **مدل (RP) Random-powerlaw**. این مدل تغییر یافته مدل ER است که توزیع درجه توانی ایجاد می‌کند [۳۱]. در این مدل نمای توزیع درجه توانی (γ) قابل تنظیم است. در تنظیماتی که برای تولید این شبکه‌ها در مجموعه داده شبکه‌های مصنوعی استفاده کردیم، پارامتر γ به صورت تصادفی از بازه $2/5 < \gamma < 3$ انتخاب شده است.

● **مدل دنیای کوچک (WS) Watts-Strogatz**. مدل کلاسیک دنیای کوچک Watts-Strogatz شبکه‌هایی با فواصل کوتاه (دنیای کوچک) و ضریب خوشه‌بندی بالا تولید می‌کند [۲۵]. این مدل از یک گراف منظم^{۱۶} آغاز می‌کند که در آن هر گره به k همسایه‌اش متصل است. سپس به صورت تصادفی برخی از یال‌ها جابجا می‌شوند. احتمال جابجایی تصادفی هر یال β خوانده می‌شود. در مجموعه مصنوعی ایجاد شده، مقدار β به صورت ثابت برابر با $\beta = 0.5$ انتخاب شده و مقدار k به صورت تصادفی در بازه اعداد صحیح بین ۲ تا ۱۰ انتخاب شده است ($2 \leq k \leq 10$).

۳.۳ شبکه‌های زمان‌دار

بخش دیگری از مجموعه داده شبکه‌های استفاده شده در این پژوهش، شامل شبکه‌هایی زمان‌دار است. هر شبکه زمان‌دار در واقع مجموعه شبکه‌هایی است که هر یک از آن‌ها با یک برچسب زمانی مشخص شده است. به عبارت دیگر، یک شبکه زمان‌دار، ساختار شبکه را به صورت یک "سری زمانی" در نظر می‌گیرد و تغییرات ساختار شبکه

^{۱۴}Shrinking Diameter

^{۱۵}Densification

^{۱۶}Regular Lattice

در گذر زمان را نشان می‌دهد. در این پژوهش، دو مجموعه شبکه زمان‌دار تهیه کردیم: اول، *CiteSeerX*، که شبکه ارجاعات بین مقالات را نشان می‌دهد و از کتابخانه دیجیتال *CiteSeer^X* استخراج شده است [۴۵] و دوم، *Collab_CiteSeerX* که شبکه همکاری در تولید مقالات را نشان می‌دهد و باز هم از *CiteSeer^X* استخراج شده است. در *CiteSeerX* (شبکه ارجاعات)، هر گره یک مقاله است و هر یال نشان از یک ارجاع از یک مقاله به یک مقاله دیگر است. در *Collab_CiteSeerX* هر گره یک نویسنده است و هر یال نشان‌دهنده وجود همکاری بین دو نویسنده برای نوشتن مقالات است. برای هر یک از این دو شبکه زمان‌دار، ۱۱ نمونه زمانی از سال‌های ۱۹۹۰ تا ۲۰۱۰، به صورت دوسالانه، تشکیل دادیم (۱۹۹۰، ۱۹۹۲، ... و ۲۰۱۰). به عنوان مثال "Cit_CiteSeerX_۱۹۹۴" شبکه ارجاعات در مقالات ثبت شده در *CiteSeer^X* تا پیش از سال ۱۹۹۴ را نشان می‌دهد.

در این پژوهش، شبکه‌های زمان‌دار از دو منظر برای ارزیابی روش‌های ابداعی استفاده شده‌اند. به عبارت دیگر، دو نوع شاهد تشابه در بین شبکه‌های زمان‌دار وجود دارد که در ارزیابی‌ها قابل استفاده هستند. شاهد اول، شباهت بین شبکه‌های هم‌نوع است. مثلاً شبکه "Cit_CiteSeerX_۲۰۱۰" بیش از آن که شبیه "Collab_CiteSeerX_۲۰۰۰" باشد شبیه شبکه "Cit_CiteSeerX_۲۰۰۰" است (زیرا هم‌نوع هستند). شاهد دوم، شباهت بین شبکه‌های هم‌نوعی است که از نظر زمانی به هم نزدیک هستند. مثلاً ساختار شبکه "Cit_CiteSeerX_۲۰۱۰" شباهت زیادی به شبکه "Cit_CiteSeerX_۲۰۰۸" دارد و شباهت کمتری به شبکه "Cit_CiteSeerX_۲۰۰۰" دارد. این شواهد شباهت، در کارهای مشابه دیگر نیز مورد اشاره یا استفاده قرار گرفته‌اند [۵۵، ۵۶].

درباره مقالات استخراج شده از رساله

مقاله‌های مجله استخراج شده از این رساله عبارتند از:

1. Feature Extraction from Degree Distribution for Comparison and Analysis of Complex Networks [85]

The Computer Journal (Oxford University Press)

این مقاله روش استخراج ویژگی از توزیع درجه را در بر می‌گیرد. مفاد این مقاله عمدتاً در فصل ۴ ارائه شده است.

2. Distance Metric Learning for Complex Networks: Towards Size-Independent Comparison of Network Structures [153]

Chaos: An Interdisciplinary Journal of Nonlinear Science

این مقاله معیار شباهت پیشنهادی برای شبکه‌های اجتماعی را گزارش می‌کند و محتوای آن در فصل ۵ گنجانده شده است.

3. Noise-Tolerant Model Selection and Parameter Estimation for Complex Networks [154]

Physica A: Statistical Mechanics and its Applications (Elsevier)

این مقاله روش انتخاب و تنظیم مدل مولد را نشان می‌دهد و در فصل ۶ گنجانده شده است.

4. Generative model selection using a scalable and size-independent complex network classifier [71]

Chaos: An Interdisciplinary Journal of Nonlinear Science

این مقاله روش اولیه پیشنهادی برای انتخاب مدل را نشان می‌دهد که در فصل ۵ توضیح داده شده است.

مقاله‌های اصلی کنفرانسی که از این رساله استخراج شده‌اند، عبارتند از:

1. A Proactive Database System and its Query Language for Social Network Simulation [155]

Conference of the European Social Simulation Association (ESSA2012)

این مقاله یک مدل مرجع اولیه را برای شبیه‌سازی مقیاس‌پذیر شبکه‌های بزرگ ارائه می‌کند که در پیوست الف به طور خلاصه توضیح داده شد. این مقاله در همایش سالانه انجمن شبیه‌سازی اجتماعی اروپا در اتریش ارائه شده است.

2. Quantification and Comparison of Degree Distributions in Complex Networks [89]

International Symposium on Telecommunications (IST2014)

این مقاله نسخه اول روش استخراج ویژگی از توزیع درجه را پوشش می‌دهد که در فصل ۴ توصیف شد.

کتاب نامه

- [1] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, “Graph structure in the web,” *Computer networks*, Vol.33, No.1, pp.309–320, 2000.
- [2] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of modern physics*, Vol.74, No.1, pp.47–97, 2002.
- [3] R. Axelrod, “Advancing the art of simulation in the social sciences,” In *Simulating social phenomena*, pp.21–40, Springer, 1997.
- [4] R. Sun. *Cognition and multi-agent interaction: From cognitive modeling to social simulation*. Cambridge University Press, 2006.
- [5] F. Squazzoni, W. Jager, and B. Edmonds, “Social simulation in the social sciences a brief overview,” *Social Science Computer Review*, Vol.32, No.3, pp.279–294, 2014.
- [6] M. J. North, C. M. Macal, J. S. Aubin, P. Thimmapuram, M. Bragen, J. Hahn, J. Karr, N. Brigham, M. E. Lacy, and D. Hampton, “Multiscale agent-based consumer market modeling,” *Complexity*, Vol.15, No.5, pp.37–47, 2010.
- [7] J. D. Farmer and D. Foley, “The economy needs agent-based modelling,” *Nature*, Vol.460, No.7256, pp.685–686, 2009.
- [8] C. Aymanns and C.-P. Georg, “Contagious synchronization and endogenous network formation in financial networks,” *Journal of Banking & Finance*, Vol.50, pp.273–285, 2015.
- [9] K. M. Carley, D. B. Fridsma, E. Casman, A. Yahja, N. Altman, L.-C. Chen, B. Kaminsky, and D. Nave, “Biowar: scalable agent-based model of bioattacks,” *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, Vol.36, No.2, pp.252–265, 2006.
- [10] J. Moffat, J. Smith, and S. Witty, “Emergent behaviour: theory and experimentation using the mana model,” *Advances in Decision Sciences*, Vol.2006, 2006.

-
- [11] C.-Y. Huang, C.-T. Sun, J.-L. Hsieh, and H. Lin, "Simulating sars: Small-world epidemiological modeling and public health policy assessments," *Journal of Artificial Societies and Social Simulation*, Vol.7, No.4, 2004.
- [12] C. Cioffi-Revilla and M. Rouleau, "Mason rebeland: An agent-based model of politics, environment, and insurgency1," *International Studies Review*, Vol.12, No.1, pp.31–52, 2010.
- [13] A. Khan, *Modeling and Augmenting "Searchable" Online Communities*. Ph.D. dissertation, Buffalo, NY, USA, 2010. AAI3423482.
- [14] M. Gatti, P. Cavalin, S. B. Neto, C. Pinhanez, C. dos Santos, D. Gribel, and A. P. Appel, "Large-scale multi-agent-based modeling and simulation of microblogging-based online social network," In *Multi-Agent-Based Simulation XIV*, pp.17–33, Springer, 2014.
- [15] E. Turban, R. Sharda, and D. Delen. *Decision Support and Business Intelligence Systems*. Upper Saddle River, NJ, USA: Prentice Hall Press, 9th Ed., 2010.
- [16] M. E. Schramm, K. J. Trainor, M. Shanker, and M. Y. Hu, "An agent-based diffusion model with consumer and brand agents," *Decision Support Systems*, Vol.50, No.1, pp.234–242, 2010.
- [17] R.-Y. Tian and Y.-J. Liu, "Isolation, insertion, and reconstruction: Three strategies to intervene in rumor spread based on supernetwork model," *Decision Support Systems*, Vol.67, pp.121–130, 2014.
- [18] C. M. Macal and M. J. North, "Tutorial on agent-based modelling and simulation," *Journal of Simulation*, Vol.4, No.3, pp.151–162, 2010.
- [19] E. Kiesling, M. Günther, C. Stummer, and L. M. Wakolbinger, "Agent-based simulation of innovation diffusion: a review," *Central European Journal of Operations Research*, Vol.20, No.2, pp.183–230, 2012.
- [20] B. Neville, M. Fasli, and J. Pitt, "Utilising social recommendation for decision-making in distributed multi-agent systems," *Expert Systems with Applications*, Vol.42, No.6, pp.2884–2906, 2015.
- [21] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Statistical properties of community structure in large social and information networks," In *Proceedings of the 17th international conference on World Wide Web*, pp.695–704, ACM, 2008.
- [22] S. Milgram, "The small world problem," *Psychology today*, Vol.2, No.1, pp.60–67, 1967.

-
- [23] R. Muhamad, *Search in Social Networks*. Ph.D. dissertation, Columbia University, 2010.
- [24] J. Leskovec and E. Horvitz, “Planetary-scale views on a large instant-messaging network,” In *Proceedings of the 17th international conference on World Wide Web*, pp.915–924, ACM, 2008.
- [25] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *nature*, Vol.393, No.6684, pp.440–442, 1998.
- [26] A. Clauset, C. R. Shalizi, and M. E. Newman, “Power-law distributions in empirical data,” *SIAM review*, Vol.51, No.4, pp.661–703, 2009.
- [27] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a highly connected world*. New York, NY, USA: Cambridge university press, 2010.
- [28] M. E. Newman, “Modularity and community structure in networks,” *Proceedings of the National Academy of Sciences of the United States of America*, Vol.103, No.23, pp.8577–8582, 2006.
- [29] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, “Kronecker graphs: An approach to modeling networks,” *The Journal of Machine Learning Research*, Vol.11, pp.985–1042, 2010.
- [30] P. Erdős and A. Rényi, “On the central limit theorem for samples from a finite population,” *Publ. Math. Inst. Hungar. Acad. Sci*, Vol.4, pp.49–61, 1959.
- [31] D. Volchenkov and P. Blanchard, “An algorithm generating random graphs with power law degree distributions,” *Physica A: Statistical Mechanics and its Applications*, Vol.315, No.3, pp.677–690, 2002.
- [32] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, Vol.286, No.5439, pp.509–512, 1999.
- [33] J. Leskovec, J. Kleinberg, and C. Faloutsos, “Graphs over time: Densification laws, shrinking diameters and possible explanations,” In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, (New York, NY, USA), pp.177–187, ACM, 2005.
- [34] M. Kim and J. Leskovec, “Multiplicative attribute graph model of real-world networks,” *Internet Mathematics*, Vol.8, No.1-2, pp.113–160, 2012.
- [35] R. Patro, G. Duggal, E. Sefer, H. Wang, D. Filippova, and C. Kingsford, “The missing models: A data-driven approach for learning how networks grow,” In *Proceedings of the 18th ACM SIGKDD International Conference*

- on Knowledge Discovery and Data Mining*, (New York, NY, USA), pp.42–50, ACM, 2012.
- [36] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, “Stochastic models for the web graph,” In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, (Washington, DC, USA), pp.57–65, IEEE Computer Society, 2000.
- [37] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins, “The web as a graph: Measurements, models, and methods,” In *Proceedings of the 5th Annual International Conference on Computing and Combinatorics*, (Berlin, Heidelberg), pp.1–17, Springer-Verlag, 1999.
- [38] L. Hamill and N. Gilbert, “Social circles: A simple structure for agent-based social network models,” *Journal of Artificial Societies and Social Simulation*, Vol.12, No.2, p.3, 2009.
- [39] M. Girvan and M. E. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, Vol.99, No.12, pp.7821–7826, 2002.
- [40] C.-T. Li and S.-D. Lin, “Social flocks: a crowd simulation framework for social network generation, community detection, and collective behavior modeling,” In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.765–768, ACM, 2011.
- [41] L. Akoglu and C. Faloutsos, “RTG: a recursive realistic graph generator using random typing,” *Data Mining and Knowledge Discovery*, Vol.19, No.2, pp.194–209, 2009.
- [42] M. Penrose. *Random geometric graphs*, Vol.5. Oxford University Press, 2003.
- [43] W. Aiello, A. Bonato, C. Cooper, J. Janssen, and P. Prałat, “A spatial web graph model with local influence regions,” *Internet Mathematics*, Vol.5, No.1-2, pp.175–196, 2008.
- [44] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. Newman, and S. H. Strogatz, “Are randomly grown graphs really random?,” *Physical Review E*, Vol.64, No.4, p.041902, 2001.
- [45] “Citeseerx digital library,” <http://citeseerx.ist.psu.edu>.
- [46] “Stanford large network dataset collection,” <http://snap.stanford.edu/data/>.
- [47] “XML repository of DBLP library,” <http://dblp.uni-trier.de/xml/>.

-
- [48] “Christian Sommer’s graph datasets,” <http://www.sommer.jp/graphs/>.
- [49] “Giulio Rossetti networks dataset,” <http://giuliorossetti.net/about/ongoing-works/datasets>.
- [50] “Alex Arenas’s network datasets,” <http://deim.urv.cat/~aarenas/data/welcome.htm>.
- [51] B. Klimt and Y. Yang, “The Enron corpus: A new dataset for email classification research,” In *Proceedings of the 15th European Conference on Machine Learning (ECML 2004)*, pp.217–226, Springer, 2004.
- [52] “The Koblenz network collection,” <http://konect.uni-koblenz.de/>.
- [53] “Newman’s NetData collection,” <http://www-personal.umich.edu/~mejn/netdata/>.
- [54] “Network datasets at Max Planck,” <http://socialnetworks.mpi-sws.org>.
- [55] P. Holme and J. Saramäki, “Temporal networks,” *Physics reports*, Vol.519, No.3, pp.97–125, 2012.
- [56] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, “Small-world behavior in time-varying graphs,” *Physical Review E*, Vol.81, No.5, p.055101, 2010.
- [57] M. E. Newman, “The structure and function of complex networks,” *SIAM review*, Vol.45, No.2, pp.167–256, 2003.
- [58] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. Villas Boas, “Characterization of complex networks: A survey of measurements,” *Advances in Physics*, Vol.56, No.1, pp.167–242, 2007.
- [59] L. Muchnik, S. Pei, L. C. Parra, S. D. Reis, J. S. Andrade Jr, S. Havlin, and H. A. Makse, “Origins of power-law degree distribution in the heterogeneity of human activity in social networks,” *Scientific reports*, Vol.3, p.1783, 2013.
- [60] J. Leskovec and C. Faloutsos, “Sampling from large graphs,” In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York, NY, USA), pp.631–636, ACM, 2006.
- [61] V. Gómez, A. Kaltenbrunner, and V. López, “Statistical analysis of the social network and discussion threads in slashdot,” In *Proceedings of the 17th International Conference on World Wide Web*, (New York, NY, USA), pp.645–654, ACM, 2008.

- [62] M. P. Stumpf, C. Wiuf, and R. M. May, "Subnets of scale-free networks are not scale-free: sampling properties of networks," *Proceedings of the National Academy of Sciences of the United States of America*, Vol.102, No.12, pp.4221–4224, 2005.
- [63] E. M. Airoidi, X. Bai, and K. M. Carley, "Network sampling and classification: An investigation of network model representations," *Decision Support Systems*, Vol.51, No.3, pp.506–518, 2011.
- [64] A. Sala, L. Cao, C. Wilson, R. Zablit, H. Zheng, and B. Y. Zhao, "Measurement-calibrated graph models for social network experiments," In *Proceedings of the 19th International Conference on World Wide Web*, (New York, NY, USA), pp.861–870, ACM, 2010.
- [65] V. Boginski, S. Butenko, and P. M. Pardalos, "Mining market data: a network approach," *Computers & Operations Research*, Vol.33, No.11, pp.3171–3184, 2006.
- [66] C. J. Stam and J. C. Reijneveld, "Graph theoretical analysis of complex networks in the brain," *Nonlinear biomedical physics*, Vol.1, No.1, p.3, 2007.
- [67] J. Leskovec, K. J. Lang, and M. Mahoney, "Empirical comparison of algorithms for network community detection," In *Proceedings of the 19th International Conference on World Wide Web*, (New York, NY, USA), pp.631–640, ACM, 2010.
- [68] M. P. Stumpf and C. Wiuf, "Sampling properties of random graphs: the degree distribution," *Physical Review E*, Vol.72, No.3, p.036118, 2005.
- [69] S. H. Lee, P.-J. Kim, and H. Jeong, "Statistical properties of sampled networks," *Physical Review E*, Vol.73, No.1, p.016102, 2006.
- [70] J.-D. J. Han, D. Dupuy, N. Bertin, M. E. Cusick, and M. Vidal, "Effect of sampling on topology predictions of protein-protein interaction networks," *Nature biotechnology*, Vol.23, No.7, pp.839–844, 2005.
- [71] S. Motallebi, S. Aliakbary, and J. Habibi, "Generative model selection using a scalable and size-independent complex network classifier," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol.23, No.4, p.043127, 2013.
- [72] J. Janssen, M. Hurshman, and N. Kalyaniwalla, "Model selection for social networks using graphlets," *Internet Mathematics*, Vol.8, No.4, pp.338–363, 2012.

- [73] M. Middendorf, E. Ziv, and C. H. Wiggins, “Inferring network mechanisms: the drosophila melanogaster protein interaction network,” *Proceedings of the National Academy of Sciences of the United States of America*, Vol.102, No.9, pp.3192–3197, 2005.
- [74] K. Juszczyszyn, N. T. Nguyen, G. Kolaczek, A. Grzech, A. Pieczynska, and R. Katarzyniak, “Agent-based approach for distributed intrusion detection system design,” In *Proceedings of the 6th International Conference on Computational Science - Volume Part III*, (Berlin, Heidelberg), pp.224–231, Springer-Verlag, 2006.
- [75] P. Papadimitriou, A. Dasdan, and H. Garcia-Molina, “Web graph similarity for anomaly detection,” *Journal of Internet Services and Applications*, Vol.1, No.1, pp.19–30, 2010.
- [76] R. Pastor-Satorras and A. Vespignani, “Epidemic dynamics in finite size scale-free networks,” *Physical Review E*, Vol.65, No.3, p.035108, 2002.
- [77] A. Montanari and A. Saberi, “The spread of innovations in social networks,” *Proceedings of the National Academy of Sciences of the United States of America*, Vol.107, No.47, pp.20196–20201, 2010.
- [78] L. Briesemeister, P. Lincoln, and P. Porras, “Epidemic profiles and defense of scale-free networks,” In *Proceedings of the 2003 ACM Workshop on Rapid Malcode*, (New York, NY, USA), pp.67–75, ACM, 2003.
- [79] G. Kossinets and D. J. Watts, “Empirical analysis of an evolving social network,” *Science*, Vol.311, No.5757, pp.88–90, 2006.
- [80] M. L. Goldstein, S. A. Morris, and G. G. Yen, “Problems with fitting to the power-law distribution,” *The European Physical Journal B-Condensed Matter and Complex Systems*, Vol.41, No.2, pp.255–258, 2004.
- [81] W. Deng, W. Li, X. Cai, and Q. A. Wang, “The exponential degree distribution in complex networks: Non-equilibrium network theory, numerical simulation and empirical data,” *Physica A: Statistical Mechanics and its Applications*, Vol.390, No.8, pp.1481–1485, 2011.
- [82] M. Faloutsos, P. Faloutsos, and C. Faloutsos, “On power-law relationships of the internet topology,” In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, (New York, NY, USA), pp.251–262, ACM, 1999.

- [83] N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, and D. Song, "Evolution of social-attribute networks: Measurements, modeling, and implications using Google+," In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, (New York, NY, USA), pp.131–144, ACM, 2012.
- [84] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?," In *Proceedings of the 19th International Conference on World Wide Web*, (New York, NY, USA), pp.591–600, ACM, 2010.
- [85] S. Aliakbary, J. Habibi, and A. Movaghar, "Feature extraction from degree distribution for comparison and analysis of complex networks," *The Computer Journal*, 2015.
- [86] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, Vol.298, No.5594, pp.824–827, 2002.
- [87] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, (New York, NY, USA), pp.29–42, ACM, 2007.
- [88] D. Corlette and F. Shipman III, "Capturing on-line social network link dynamics using event-driven sampling," In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04*, (Washington, DC, USA), pp.284–291, IEEE Computer Society, 2009.
- [89] S. Aliakbary, J. Habibi, and A. Movaghar, "Quantification and comparison of degree distributions in complex networks," In *Proceedings of The Seventh International Symposium on Telecommunications (IST2014)*, pp.464–469, IEEE, IEEE, 2014.
- [90] J. P. Bagrow, E. M. Bollt, J. D. Skufca, and D. Ben-Avraham, "Portraits of complex networks," *EPL (Europhysics Letters)*, Vol.81, No.6, p.68004, 2008.
- [91] M. Berlingerio, D. Koutra, T. Eliassi-Rad, and C. Faloutsos, "Network similarity via multiple social theories," In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, (New York, NY, USA), pp.1439–1440, ACM, 2013.
- [92] A. Mehler, "Structural similarities of complex networks: A computational model by example of wiki graphs," *Applied Artificial Intelligence*, Vol.22, No.7-8, pp.619–683, 2008.

- [93] J.-P. Onnela, D. J. Fenn, S. Reid, M. A. Porter, P. J. Mucha, M. D. Fricker, and N. S. Jones, "Taxonomies of networks from community structure," *Physical Review E*, Vol.86, No.3, p.036104, 2012.
- [94] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, Vol.13, No.1, pp.21–27, 1967.
- [95] C.-N. J. Yu and T. Joachims, "Learning structural svms with latent variables," In *Proceedings of the 26th Annual International Conference on Machine Learning*, (New York, NY, USA), pp.1169–1176, ACM, 2009.
- [96] J. Bian, Y. Liu, E. Agichtein, and H. Zha, "Finding the right facts in the crowd: Factoid question answering over social media," In *Proceedings of the 17th International Conference on World Wide Web*, (New York, NY, USA), pp.467–476, ACM, 2008.
- [97] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, Vol.28, No.3, pp.301–315, 1998.
- [98] M. E. Newman, "Assortative mixing in networks," *Physical review letters*, Vol.89, No.20, p.208701, 2002.
- [99] J. C. Platt, "Advances in kernel methods," Ch. Fast Training of Support Vector Machines Using Sequential Minimal Optimization, pp.185–208, Cambridge, MA, USA: MIT Press, 1999.
- [100] N. Pržulj, D. G. Corneil, and I. Jurisica, "Modeling interactome: scale-free or geometric?," *Bioinformatics*, Vol.20, No.18, pp.3508–3515, 2004.
- [101] D. Koutra, J. T. Vogelstein, and C. Faloutsos, "DELTACON: A principled massive-graph similarity function," In *Proceedings of the 13th SIAM International Conference on Data Mining (SDM)*, pp.162–170, 2013.
- [102] N. Pržulj, "Biological network comparison using graphlet degree distribution," *Bioinformatics*, Vol.23, No.2, pp.e177–e183, 2007.
- [103] S. Soundarajan, T. Eliassi-Rad, and B. Gallagher, "A guide to selecting a network similarity method," In *Proceedings of the 2014 SIAM International Conference on Data Mining*, SIAM, 2014.
- [104] Ö. N. Yaveroglu, N. Malod-Dognin, D. Davis, Z. Levnajic, V. Janjic, R. Karapandza, A. Stojmirovic, and N. Pržulj, "Revealing the hidden language of complex networks," *Scientific reports*, Vol.4, 2014.

-
- [105] R. C. Wilson and P. Zhu, "A study of graph spectra for comparing graphs and trees," *Pattern Recognition*, Vol.41, No.9, pp.2833–2841, 2008.
- [106] L. A. Zager and G. C. Verghese, "Graph similarity scoring and matching," *Applied mathematics letters*, Vol.21, No.1, pp.86–94, 2008.
- [107] S. Vishwanathan, N. N. Schraudolph, R. Kondor, and K. M. Borgwardt, "Graph kernels," *The Journal of Machine Learning Research*, Vol.99, pp.1201–1242, 2010.
- [108] H. Kashima, K. Tsuda, and A. Inokuchi, "Kernels for graphs," *Kernel methods in computational biology*, Vol.39, No.1, pp.101–113, 2004.
- [109] T. Gärtner, P. Flach, and S. Wrobel, "On graph kernels: Hardness results and efficient alternatives," In *Learning Theory and Kernel Machines*, pp.129–143, Springer, 2003.
- [110] U. Kang, H. Tong, and J. Sun, "Fast random walk graph kernel.," In *SDM*, pp.828–838, 2012.
- [111] K. M. Borgwardt, *Graph kernels*. Ph.D. dissertation, Ludwig-Maximilians-Universität München, 2007.
- [112] A. Kelmans, "Comparison of graphs by their number of spanning trees," *Discrete Mathematics*, Vol.16, No.3, pp.241–261, 1976.
- [113] K. Faust, "Comparing social networks: size, density, and local structure," *Metodološki zvezki*, Vol.3, No.2, pp.185–216, 2006.
- [114] I. Bordino, D. Donato, A. Gionis, and S. Leonardi, "Mining large networks with subgraph counting," In *Eighth IEEE International Conference on Data Mining (ICDM'08)*, pp.737–742, IEEE, 2008.
- [115] R. Kondor, N. Shervashidze, and K. M. Borgwardt, "The graphlet spectrum," In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp.529–536, ACM, 2009.
- [116] N. Shervashidze, T. Petri, K. Mehlhorn, K. M. Borgwardt, and S. Vishwanathan, "Efficient graphlet kernels for large graph comparison," In *International Conference on Artificial Intelligence and Statistics*, pp.488–495, 2009.
- [117] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic topology analysis and generation using degree correlations," *ACM SIGCOMM Computer Communication Review*, Vol.36, No.4, pp.135–146, 2006.

-
- [118] L. Yang and R. Jin, “Distance metric learning: A comprehensive survey,” *Michigan State University*, pp.1–51, 2006.
- [119] E. P. Xing, M. I. Jordan, S. Russell, and A. Ng, “Distance metric learning with application to clustering with side-information,” In *Advances in neural information processing systems*, pp.505–512, 2002.
- [120] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *The Journal of Machine Learning Research*, Vol.10, pp.207–244, 2009.
- [121] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1st Ed., 1989.
- [122] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, “Controllability of complex networks,” *Nature*, Vol.473, No.7346, pp.167–173, 2011.
- [123] V. Latora and M. Marchiori, “Vulnerability and protection of infrastructure networks,” *Physical Review E*, Vol.71, No.1, p.015103, 2005.
- [124] J. M. Kleinberg, “Navigation in a small world,” *Nature*, Vol.406, No.6798, pp.845–845, 2000.
- [125] M. Rahman, M. Bhuiyan, and M. A. Hasan, “Graft: an approximate graphlet counting algorithm for large graph analysis,” In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp.1467–1471, ACM, 2012.
- [126] J. C. Dunn, “A fuzzy relative of the isodata process and its use in detection compact well-separated clusters,” *Journal of Cybernetics*, Vol.3, pp.32–57, 1973.
- [127] M. Fazli, M. Ghodsi, J. Habibi, P. Jalaly Khalilabadi, V. Mirrokni, and S. S. Sadeghabad, “On the non-progressive spread of influence through social networks,” In *Proceedings of the 10th Latin American International Conference on Theoretical Informatics, LATIN’12*, (Berlin, Heidelberg), pp.315–326, Springer-Verlag, 2012.
- [128] E. Le Martelot and C. Hankin, “Fast multi-scale detection of relevant communities in large-scale networks,” *The Computer Journal*, Vol.56, No.9, pp.1136–1150, 2013.
- [129] J. Leskovec and R. Sosič, “SNAP: A general purpose network analysis and graph mining library in C++,” <http://snap.stanford.edu/snap>, June 2014.

-
- [130] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal, Complex Systems*, Vol.1695, No.5, 2006.
- [131] K. Q. Weinberger, "LMNN software tool for large margin nearest neighbors," <http://www.cse.wustl.edu/~kilian/code/code.html>, 2014.
- [132] O. Kuchaiev, A. Stevanović, W. Hayes, and N. Pržulj, "Graphcrunch 2: software tool for network modeling, alignment and clustering," *BMC bioinformatics*, Vol.12, p.24, 2011.
- [133] S. Chester, B. M. Kapron, G. Srivastava, and S. Venkatesh, "Complexity of social network anonymization," *Social Network Analysis and Mining*, Vol.3, No.2, pp.151–166, 2013.
- [134] B. Zhou, J. Pei, and W. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," *ACM SIGKDD Explorations Newsletter*, Vol.10, No.2, pp.12–22, 2008.
- [135] G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A practical attack to de-anonymize social network users," In *Security and Privacy (SP), 2010 IEEE Symposium on*, pp.223–238, IEEE, 2010.
- [136] S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava, "Class-based graph anonymization for social network data," *Proceedings of the VLDB Endowment*, Vol.2, No.1, pp.766–777, 2009.
- [137] S. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica*, Vol.31, pp.249–268, 2007.
- [138] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," In *Advances in kernel methods*, pp.185–208, MIT press, 1999.
- [139] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, Vol.2, No.5, pp.359–366, 1989.
- [140] G. Holmes, B. Pfahringer, R. Kirkby, E. Frank, and M. Hall, "Multiclass alternating decision trees," In *Proceedings of the 13th European Conference on Machine Learning, ECML '02*, (London, UK, UK), pp.161–172, Springer-Verlag, 2002.
- [141] S. Motallebi, "Model selection for complex network generation," Master's thesis, Sharif University of Technology, 2013. Persian.
- [142] M. Alam, M. Khan, and M. V. Marathe, "Distributed-memory parallel algorithms for generating massive scale-free networks using preferential attachment model," In *Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis*, p.91, ACM, 2013.

- [143] V. Batagelj and U. Brandes, “Efficient generation of large random networks,” *Physical Review E*, Vol.71, No.3, p.036113, 2005.
- [144] S. Tisue and U. Wilensky, “NetLogo: Design and implementation of a multi-agent modeling environment,” In *The Agent 2004 Conference on: Social Dynamics: Interaction, Reflexivity, and Emergence*, pp.7–9, 2004.
- [145] N. Collier, “Repast: An extensible framework for agent simulation,” *The University of Chicago’s Social Science Research*, Vol.36, p.2003, 2003.
- [146] “Swarm development group,” <http://www.swarm.org>.
- [147] S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan, “Mason: A multiagent simulation environment,” *Simulation*, Vol.81, No.7, pp.517–527, 2005.
- [148] S. Luke, “MASON multiagent simulation toolkit,” <http://cs.gmu.edu/~eclab/projects/mason/>, 2014.
- [149] X. Li, W. Mao, D. Zeng, and F.-Y. Wang, “Agent-based social simulation and modeling in social computing,” *Intelligence and Security Informatics*, pp.401–412, 2008.
- [150] R. Ronen and O. Shmueli, “SoQL: A language for querying and creating data in social networks,” In *Data Engineering, 2009. ICDE’09. IEEE 25th International Conference on*, pp.1595–1602, IEEE, 2009.
- [151] A. Dries, S. Nijssen, and L. De Raedt, “BiQL: a query language for analyzing information networks,” In *Bisociative Knowledge Discovery*, pp.147–165, Springer, 2012.
- [152] “Facebook Query Language (FQL),” <https://developers.facebook.com/docs/reference/fql/>.
- [153] S. Aliakbary, S. Motallebi, S. Rashidian, J. Habibi, and A. Movaghar, “Distance metric learning for complex networks: Towards size-independent comparison of network structures,” *Chaos: An Interdisciplinary Journal of Non-linear Science*, Vol.25, p.023111, 2015.
- [154] S. Aliakbary, S. Motallebi, S. Rashidian, J. Habibi, and A. Movaghar, “Noise-tolerant model selection and parameter estimation for complex networks,” *Physica A: Statistical Mechanics and its Applications*, Vol.427, p.100–112, 2015.

- [155] S. Aliakbary and J. Habibi, “A proactive database system and its query language for social network simulation,” In *Proceedings of Eighth Conference of the European Social Simulation Association (ESSA)*, pp.255–264, 2012.

واژه‌نامه انگلیسی به فارسی

Agent	عامل
Agent-based	مبتهی بر عامل
Approximator	تخمین‌زننده
Artificial Networks	شبکه‌های مصنوعی
Artificial neural networks	شبکه‌های عصبی مصنوعی
Assortativity	جورش
Average clustering coefficient	میانگین ضریب خوشه‌بندی
Average Degree	میانگین درجه
Average path length	میانگین فاصله
Baseline methods	روش‌های پایه
Biased Results	سوگیری نتایج
Canberra	کانبرا
Citation Networks	شبکه‌های ارجاعات
Class	رده
Classification	رده‌بندی
Clustering Coefficient	ضریب خوشه‌بندی
Cognitive Science	علوم شناختی
Collaboration Networks	شبکه‌های همکاری
Communication Networks	شبکه‌های ارتباطات
Community Detection	کشف انجمن‌ها
Complex Networks	شبکه‌های پیچیده
Constrained convex programming	برنامه‌سازی محدب محدود
Controllability	کنترل‌پذیری
Coordinator	هماهنگ‌ساز
Crawler	خزشگر
Cross Validation	اعتبارسنجی گردش
Crossover	مقاطع
Cumulative Distribution Function (CDF)	تابع توزیع تجمعی
Cycle	چرخه
Dataset	مجموعه داده
Decision Support Systems	سیستم‌های تصمیم‌یار
Degree Distribution	توزیع درجه
Dense	چگال
Densification	چگال‌شدن

Density	چگالی
Diameter	قطر
Discrete-time simulation	شبیه‌سازی زمان گسسته
Distance Metric	تابع فاصله
Distance Metric Learning	یادگیری تابع فاصله
Distribution Percentiles	صدک‌های توزیع
DSS	سیستم‌های تصمیم‌یار
Dunn Index	شاخص دان
Edge	یال
Edit distance	فاصله ویرایش
Effective diameter	قطر مؤثر
Efficiency	کارایی
Eigenvalue	مقدار ویژه
Euclidean	اقلیدسی
Feature Selection	انتخاب ویژگی
Feature Vector	بردار ویژگی
Fitness	برازش
Fitness Function	تابع برازش
Fitted Power-law exponent	نمای توزیع توانی برازش شده
Fixed-Length	اندازه ثابت
Frequent subgraph mining	کاوش زیرگراف‌های رایج
Friendship Networks	شبکه‌های دوستی
Generative Model	مدل مولد
Global features	ویژگی‌های سراسری
Graph	گراف
Graph alignment	ترازایی گراف
Graph isomorphism	یک‌ریختی گراف
Graph kernel	هسته گراف
Graph matching	جورایی گراف‌ها
graph summarization	خلاصه‌سازی گراف
Graphlet Counting	شمارش گرافلت
Homophily	هم‌رنگی
Initiator matrix	ماتریس اولیه
Inter-class distances	فواصل بین‌گروهی
Interval	بازه
Interval Degree Probability	احتمال بازه درجه
Intra-class distances	فواصل درون‌گروهی
Isomorph	یک‌ریخت
Iteration	تکرار
Kernel Methods	روش‌های هسته
Kolmogorov-Smirnov test (KS-test)	آزمون کولموگوروف-اسمیرنوف

Labeled	برچسب‌دار
Lazy Learning	یادگیری تنبل
local features	ویژگی‌های محلی
Log-normal distribution	توزیع لاگ-نرمال
long-tailed degree distribution	توزیع درجه دم‌دراز
Mahalanobis	ماهالانوبیس
Manhattan	منهتن
Manual	دستی
Manual Selection	انتخاب دستی
Maximum Likelihood Estimation (MLE)	برآورد درست‌نمایی بیشینه
Mean Square Error	خطای میانگین مربعات
Model Fitting	برازش مدل
Modularity	پیمانگی
Mutation	جهش
Natural Selection	انتخاب طبیعی
Navigability	ناوش‌پذیری
Nearest-Neighbor Classifier	رده‌بند مبتنی بر نزدیک‌ترین همسایه
Network	شبکه
Network anonymization	گمنام‌سازی شبکه
Network Generative Model	مدل مولد شبکه
Network Resilience	حالت ارتجاعی شبکه
Network Science	علم شبکه
Node	گره
Noise Tolerance	مقاومت به نویز
Optimizer	بهینه‌ساز
Outlier	برون‌نهاد
Passive	منفعل
Pearson Correlation	همبستگی پیرسون
Pearson Correlation Coefficient	ضریب همبستگی پیرسون
Peer-to-Peer Networks	شبکه‌های نظیربه‌نظیر
Perceptron	پرسپترون
Permission Checker	ممیز دسترسی‌ها
Positive Definite	مثبت-معین
Power-law degree distribution	توزیع درجه توانی
power-law exponent	نمای توزیع توانی
Precision at K	دقت در K
Preferential Attachment	اتصال ترجیحی
Proactive	پیش‌نگر
Proactive Database	پایگاه داده پیش‌نگر
Quantification	کمّی‌سازی
Query Executor	اجراگر پرسمان

Query Language.....	زبان پرسمان
Query Processor.....	پردازشگر پرسمان
Real Networks.....	شبکه‌های حقیقی
Region.....	ناحیه
Regular Lattice.....	گراف منظم
Rich-Club.....	ریچ کلاب
Sample.....	نمونه
Sampling.....	نمونه برداری
Scale free.....	مستقل از مقیاس
Scheduler.....	زمان‌بند
Sequential Minimal Optimization.....	بهینه‌سازی دنباله‌ای کمینه
Shrinking Diameter.....	کاهش قطر
Signature.....	امضا
Similarity Measure.....	معیار شباهت
Simulation.....	شبیه‌سازی
Small World.....	دنیای کوچک
Snapshot.....	برهه
Social Networks.....	شبکه‌های اجتماعی
Social Networks Simulation.....	شبیه‌سازی شبکه‌های اجتماعی
Social Simulation.....	شبیه‌سازی اجتماعی
Sparse.....	تُنک
Sub-iteration.....	زیرتکرار
Supervised Machine Learning.....	یادگیری ماشین با نظارت
Target network.....	شبکه هدف
Temporal Networks.....	شبکه‌های زمان‌دار
Thread.....	بند
Topological Descriptor.....	توصیف‌گر ساختار
Transitivity.....	تراگذری
Trigger.....	ماشه
Unsupervised Machine Learning.....	یادگیری ماشین بدون نظارت
Vulnerability.....	آسیب‌پذیری
What-if Scenario.....	سناریوی چه می‌شود اگر
Web Graphs.....	گراف‌های وب

واژه‌نامه فارسی به انگلیسی

Preferential Attachment	اتصال ترجیحی
Query Executor	اجراگر پرسمان
Interval Degree Probability	احتمال بازه درجه
Cross Validation	اعتبارسنجی گردشی
Euclidean	اقلیدسی
Signature	امضا
Manual Selection	انتخاب دستی
Natural Selection	انتخاب طبیعی
Feature Selection	انتخاب ویژگی
Fixed-Length	اندازه ثابت
Kolmogorov-Smirnov test (KS-test)	آزمون کولموگوروف-اسمیرنوف
Vulnerability	آسیب‌پذیری
Interval	بازه
Fitness	برازش
Model Fitting	برازش مدل
Maximum Likelihood Estimation (MLE)	برآورد درست‌نمایی بیشینه
Labeled	برچسب‌دار
Feature Vector	بردار ویژگی
Constrained convex programming	برنامه‌سازی محدب محدود
Snapshot	برهه
Outlier	برون‌نهاد
Thread	بند
Optimizer	بهینه‌ساز
Sequential Minimal Optimization	بهینه‌سازی دنباله‌ای کمینه
Proactive Database	پایگاه داده پیش‌نگر
Query Processor	پردازشگر پرسمان
Perceptron	پرسپترون
Proactive	پیش‌نگر
Sparse	تنک
Fitness Function	تابع برازش
Cumulative Distribution Function (CDF)	تابع توزیع تجمعی
Distance Metric	تابع فاصله
Approximator	تخمین‌زننده
Graph alignment	ترازیابی گراف

Transitivity	تراگذری
Iteration	تکرار
Degree Distribution	توزیع درجه
Power-law degree distribution	توزیع درجه توانی
long-tailed degree distribution	توزیع درجه دم‌دراز
Log-normal distribution	توزیع لاگ-نرمال
Topological Descriptor	توصیف‌گر ساختار
Mutation	جهش
Assortativity	جورش
Graph matching	جوریابی گراف‌ها
Cycle	چرخه
Dense	چگال
Densification	چگال‌شدن
Density	چگالی
Network Resilience	حالت ارتجاعی شبکه
Crawler	خزشگر
Mean Square Error	خطای میانگین مربعات
graph summarization	خلاصه‌سازی گراف
Manual	دستی
Precision at K	دقت در K
Small World	دنیای کوچک
Class	رده
Nearest-Neighbor Classifier	رده‌بند مبتنی بر نزدیک‌ترین همسایه
Classification	رده‌بندی
Baseline methods	روش‌های پایه
Kernel Methods	روش‌های هسته
Rich-Club	ریچ کلاب
Query Language	زبان پرسمان
Scheduler	زمان‌بند
Sub-iteration	زیرتکرار
What-if Scenario	سناریوی چه می‌شود اگر
Biased Results	سوگیری نتایج
Decision Support Systems	سیستم‌های تصمیم‌یار
DSS	سیستم‌های تصمیم‌یار
Dunn Index	شاخص دان
Network	شبکه
Target network	شبکه هدف
Social Networks	شبکه‌های اجتماعی
Communication Networks	شبکه‌های ارتباطات
Citation Networks	شبکه‌های ارجاعات
Complex Networks	شبکه‌های پیچیده

Real Networks	شبکه‌های حقیقی
Friendship Networks	شبکه‌های دوستی
Temporal Networks	شبکه‌های زمان‌دار
Artificial neural networks	شبکه‌های عصبی مصنوعی
Artificial Networks	شبکه‌های مصنوعی
Peer-to-Peer Networks	شبکه‌های نظریه‌نظیر
Collaboration Networks	شبکه‌های همکاری
Simulation	شبیه‌سازی
Social Simulation	شبیه‌سازی اجتماعی
Discrete-time simulation	شبیه‌سازی زمان گسسته
Social Networks Simulation	شبیه‌سازی شبکه‌های اجتماعی
Graphlet Counting	شمارش گرافلت
Distribution Percentiles	صدک‌های توزیع
Clustering Coefficient	ضریب خوشه‌بندی
Pearson Correlation Coefficient	ضریب همبستگی پیرسون
Agent	عامل
Network Science	علم شبکه
Cognitive Science	علوم شناختی
Edit distance	فاصله ویرایش
Inter-class distances	فواصل بین گروهی
Intra-class distances	فواصل درون گروهی
Diameter	قطر
Effective diameter	قطر مؤثر
Efficiency	کارایی
Canberra	کانبرا
Shrinking Diameter	کاهش قطر
Frequent subgraph mining	کاوش زیرگراف‌های رایج
Community Detection	کشف انجمن‌ها
Quantification	کمّی‌سازی
Controllability	کنترل‌پذیری
Graph	گراف
Regular Lattice	گراف منظم
Web Graphs	گراف‌های وب
Node	گره
Network anonymization	گمنام‌سازی شبکه
Initiator matrix	ماتریس اولیه
Modularity	پیمانگی
Trigger	ماشه
Mahalanobis	ماهالانوبیس
Agent-based	مبتنی بر عامل
Crossover	مقاطع

Positive Definite	مثبت-معین
Dataset	مجموعه داده
Generative Model	مدل مولد
Network Generative Model	مدل مولد شبکه
Scale free	مستقل از مقیاس
Similarity Measure	معیار شباهت
Noise Tolerance	مقاومت به نویز
Eigenvalue	مقدار ویژه
Permission Checker	ممیز دسترسی‌ها
Passive	منفعل
Manhattan	منهتن
Average Degree	میانگین درجه
Average clustering coefficient	میانگین ضریب خوشه‌بندی
Average path length	میانگین فاصله
Region	ناحیه
Navigability	ناوش‌پذیری
power-law exponent	نمای توزیع توانی
Fitted Power-law exponent	نمای توزیع توانی برازش شده
Sample	نمونه
Sampling	نمونه‌برداری
Graph kernel	هسته گراف
Coordinator	هماهنگ‌ساز
Pearson Correlation	همبستگی پیرسون
Homophily	هم‌رنگی
Global features	ویژگی‌های سراسری
local features	ویژگی‌های محلی
Distance Metric Learning	یادگیری تابع فاصله
Lazy Learning	یادگیری تنبل
Supervised Machine Learning	یادگیری ماشین با نظارت
Unsupervised Machine Learning	یادگیری ماشین بدون نظارت
Edge	یال
Isomorph	یک‌ریخت
Graph isomorphism	یک‌ریختی گراف

Abstract

A social network represents a set of entities and their relationships. Telecommunication networks, online social networks, and paper citation networks are some examples of networks in real world. Nowadays, analysis of social networks is an interesting research area with important applications. Particularly, managers of the social networks and the decision makers often require intelligent decision support for futures study in these social systems. The demanded decision support systems make it possible to define the desired social problem and to analyze the "what-if scenarios." Computer simulation is an appropriate approach toward such decision support systems. In this approach, the desired network is simulated, different scenarios are examined, and the results are studied. Simulation is an effective method for analysis of theories and to propose new hypotheses for social systems. The aim of this dissertation is to investigate open research problems towards a decision support system based on simulation of social networks. In the desired system, the decision maker can define the properties of the target network, synthesize an artificial network with the desired size and properties, run different social scenarios in the synthesized networks, and study the results of the simulations.

Generation of customizable artificial networks is one of the main research problems in this dissertation. Artificial networks are stochastic graphs which are structurally similar to the target network. In this dissertation, we proposed a framework for generation of social networks based on some research steps. First, we proposed a novel method for feature extraction from the degree distribution of the social networks. Then, we proposed a similarity metric for comparing structure of social networks. The proposed similarity metric is utilized in our next step of research which involves generative model selection. Model selection is the art of finding the best generative model, among the candidates, which is capable of generating networks similar to the target network. The proposed model fitting method not only selects the best fitting model, but also tunes the model parameters in order to generate the best fitting network with the desired size. The proposed model fitting method is robust to noise. All of the proposed methods, which are based on machine learning, outperform the existing baselines with respect to accuracy and efficiency. In this dissertation, we have also presented an initial software reference model for scalable simulation of large social networks.

Keywords: *Social Networks, Social Simulation, Decision Support Systems, Network Generative Model, Data Mining, Complex Network, Agent-based Modeling*



Sharif University of Technology

PhD Dissertation

Topic

Model Selection for Social Network
Simulation in a Decision Support
System

By

Sadegh Aliakbary

Supervisor

Dr. Jafar Habibi and Prof. Ali Movaghar

February 2015