

## سوال ۱:

در نظریه  $\epsilon$ -decaying برای یک روش ابتکاری (Heuristic) در پیش‌بینی لینک، چگونه معیار میزان "اهمیت" یا "وزن" هر یک از مسیرهای دورتر در نمودار کاهش می‌یابد و این کاهش وزن تا چه حد تضمین می‌کند که زیرگراف‌های  $h$ -تایی (h-Hop) بتوانند اطلاعات کافی برای تقریب شاخص‌های مرتبه‌ی بالا (High-Order) فراهم آورند؟ همچنین، در چه شرایطی این تقریب ممکن است دچار افت شود؟

پاسخ:

طبق نظریه  $\epsilon$ -decaying، اغلب شاخص‌های مرتبه‌ی بالا (مانند Katz، PageRank، و SimRank) به صورت جمع موزون مسیرها یا رندم‌واک‌ها فرمول‌بندی می‌شوند، که وزن آن‌ها با ضریبی کوچک‌تر از ۱ (مثلاً ضریب تضعیف  $\alpha$  یا  $\beta$  به صورت نمایی کاهش می‌یابد. این ضریب کوچک باعث می‌شود مسیرهای دورتر به شکل محسوسی کم‌اثر شوند و در نتیجه زیرگراف‌های محلی کوچک) در شعاع  $h$  از دو گره هدف (حاوی بخش عمده‌ای از اطلاعات مهم باشند. به عبارت دیگر، وقتی  $\epsilon$  کوچک باشد و ضرایب کاهشی (decay factors) به اندازه کافی کوچک‌تر از ۱ انتخاب شوند، سهم مسیرهای طولانی بر خروجی نهایی ناچیز خواهد بود. از این رو، حتی با داشتن زیرگراف‌های  $h$ -تایی، می‌توان چنان مسیرهایی را با خطای بسیار پایین تقریب زد. با این حال، اگر پارامتر تضعیف به درستی تنظیم نشود یا شبکه دارای ساختارهایی باشد که تأثیر مسیرهای دور دست به طور غیرعادی بالا باقی بماند (مثلاً وجود گره‌های ابر-متصل با درجه‌ی بسیار زیاد که باعث حضور مسیرهای متعدد و کوتاه به مناطق دورتر می‌شوند)، ممکن است حدود خطای تقریب افزایش پیدا کند. در چنین شرایطی، شعاع  $h$  کوچک دیگر برای دستیابی به خطای کم کافی نخواهد بود و می‌بایست از  $h$  بزرگ‌تری استفاده کرد یا از روش‌های جبرانی (نظیر نمونه‌گیری تصادفی) بهره برد.

## سوال ۲:

در چارچوب SEAL، برچسب‌گذاری (Labeling) گره‌ها در مرحله‌ی «ساخت ماتریس اطلاعات گره‌ها» چه نقشی در تمایز وظایف گره‌های مرکز (Center Nodes) و گره‌های مجاور ایفا می‌کند و چگونه نبود یک طرح برچسب‌گذاری مناسب می‌تواند دقت مدل گراف‌عصبی (GNN) را کاهش دهد؟

پاسخ:

در روش SEAL، از الگوریتم Labeling دو-شعاعی (Double-Radius Node Labeling, DRNL) برای تمایز جایگاه هر گره نسبت به دو گره‌ی مرکزی (که بین آن‌ها احتمال لینک را پیش‌بینی می‌کنیم) بهره گرفته می‌شود. در این برچسب‌گذاری، هر گره بر اساس فاصله‌اش از دو گره‌ی مرکز، یک برچسب عددی متمایز دریافت می‌کند تا مدل متوجه شود کدام گره‌ها همان «مبدأ/مقصد» احتمالی لینک هستند و کدام گره‌ها صرفاً همسایه یا گره جانبی می‌باشند. این تفاوت برچسب باعث می‌شود شبکه عصبی گرافی بتواند نقش ساختاری گره‌ها را تفکیک کند. اگر چنین نشانه‌گذاری صریحی وجود نداشته باشد، همه‌ی گره‌ها از دید GNN هم‌ارزش تلقی خواهند شد و شبکه دشوارتر می‌تواند تشخیص دهد کدام گره‌ها مهم‌تر هستند (مثلاً مبدأ/مقصد لینک) یا کدام گره‌ها صرفاً در حاشیه‌ی زیرگراف قرار دارند. به این ترتیب، از بین رفتن این تفکیک ساختاری، سبب می‌شود مدل در

«مکان‌یابی» صحیح الگوی اتصال (Connectivity Pattern) دچار ابهام شده و دقت پیش‌بینی لینک کاهش قابل‌توجهی پیدا کند.

### سوال ۳:

با توجه به معماری کلی SEAL که از سه بخش اصلی «استخراج زیرگراف‌های محلی»، «ساخت ماتریس اطلاعات (برچسب‌ها، بردارهای نهفته، ویژگی‌های آشکار)» و «یادگیری با شبکه عصبی گرافی» تشکیل شده است، کدام مرحله از نظر پیچیدگی زمانی بالقوه دشوارتر است و چه راهکارهایی برای کنترل اندازه یا نمونه‌گیری زیرگراف‌ها در شبکه‌های بسیار بزرگ پیشنهاد می‌شود؟

#### پاسخ:

در فرایند SEAL، مرحله‌ی استخراج زیرگراف‌های محلی (Enclosing Subgraphs) در اطراف هر زوج گره بالقوه برای لینک، معمولاً پراستفاده‌ترین منبع زمان و حافظه است. زیرا اگر گره‌های مرکزی درجه‌ی بالایی داشته باشند، زیرگراف محلی می‌تواند به سرعت بزرگ شود و به‌ویژه در شعاع‌های بالاتر (مثلاً  $h=3$  به بالا) (اندازه‌ی زیرگراف می‌تواند انفجاری شود و به‌ویژه در شبکه‌های بزرگ زمان‌بر خواهد بود. برای حل این مشکل، می‌توان از روش‌های زیر استفاده کرد: ۱. برش درجه یا فیلتر کردن همسایه‌ها: از میان همسایه‌های خیلی پرشمار یک گره، تنها مجموعه‌ای نمونه‌گیری شود تا اندازه‌ی زیرگراف کنترل گردد. ۲. انتخاب تصادفی زیرمجموعه‌ای از مسیرها: اگر ساختار گراف چگال باشد، از نمونه‌گیری تصادفی یال‌ها یا مسیرها استفاده می‌شود تا زیرگراف در حد معقول باقی بماند. ۳. تعیین حد آستانه (Threshold) برای  $h$ : بنا بر تحلیل تئوری  $\epsilon$ -decaying، مقدار  $h$  معمولاً کوچک (۱ یا ۲) در نظر گرفته می‌شود و به‌لحاظ عملی نیز نتایج خوبی می‌دهد و مشکل انفجار اندازه‌ی زیرگراف را کمتر می‌کند.

### سوال ۴:

روش SEAL چگونه می‌تواند در شبکه‌هایی که علاوه بر ساختار لینک‌ها، دارای ویژگی‌های گره‌ای (Attributes) یا بردارهای نهفته (Embeddings) هستند، از اطلاعات اضافی بهره‌برداری کند و چه تمهیدی برای جلوگیری از «یادگیری ساده» لینک‌های مثبت و منفی در بردارهای نهفته پیش‌بینی شده است؟

#### پاسخ:

چارچوب SEAL به‌صورت ذاتی می‌تواند ویژگی‌های مختلفی را در ماتریس اطلاعات گره‌ها ادغام کند. یعنی علاوه بر برچسب‌های ساختاری، بردارهای نهفته (مثلاً حاصل از node2vec یا روش‌های دیگر (و ویژگی‌های واقعی گره (attributes) را نیز ضمیمه ورودی GNN می‌کند. این کار باعث می‌شود مدل به‌صورت همزمان هم «ویژگی‌های مستقیم ساختار زیرگراف» را ببیند و هم «ویژگی‌های کم‌وبیش جهانی» هر گره (مانند embedding یا ویژگی آشکار). (یک چالش مهم این است که اگر بردارهای نهفته روی گراف اصلی آموزش داده شوند، لینک‌های واقعی (مثبت) به‌صورت مستقیم در آن embedding منعکس شده و ممکن است مدل تنها با نگاه به این بردارها متوجه وجود لینک شود و نیازی به یادگیری الگوهای ساختاری نداشته باشد. برای حل این مشکل، نویسندگان مقاله روش Negative Injection را پیشنهاد داده‌اند. در این روش، لینک‌های منفی (زوج‌های گره‌هایی که لینک ندارند) به‌طور موقت به گراف اضافه می‌شوند تا embeddings نتوانند صرفاً از وجود یا نبود یال واقعی اطلاع بیابند. بدین ترتیب،

embeddings برای تمام لینک‌های مثبت و منفی اطلاعات مشابهی ثبت می‌کنند و قابلیت تمایزگذاری ساده از بین می‌رود. در نتیجه، GNN وادار می‌شود اطلاعات ساختاری عمیق‌تری را از زیرگراف‌ها بیاموزد و عملکرد مدل بهبود یابد.