**ساختارهای گرافی از داده‌کاو‌ها: رویکردی مبتنی بر نظریه گراف و پردازش سیگنال گرافی**

مقدمه و انگیزه  
بسیاری از داده‌های واقعی، به طور ذاتی شامل ساختارهای مخفی هستند که به صورت طبیعی می‌توان آنها را به عنوان گراف نمایش داد. اما این ساختارها اغلب مستقیماً قابل دیدن نیستند و باید استنتاج شوند. ساختن یک گراف مناسب از داده‌ها، امکان استفاده از ابزارهای قدرتمند نظریه گراف و پردازش سیگنال گرافی (GSP) را برای وظایفی مانند خوشه‌بندی، کاهش ابعاد، نویززدایی سیگنال و آموزش نیمه‌نظارتی فراهم می‌کند.  
هدف این تحقیق، کاوش سیستماتیک روش‌هایی برای ساخت گراف‌های Directed یا Undirected از داده‌ها است، که بر اصول نظریه گراف و پیش‌فرض‌های نرمی سیگنال از GSP تکیه دارند. این کار پل ارتباطی بین داده‌های خام و مدل‌سازی مبتنی بر گراف را ایجاد می‌کند و توانایی استخراج بینش‌های معنادار از داده‌های پیچیده را تقویت می‌نماید.

پیشینه و کارهای مرتبط  
جریان‌های پژوهشی مختلفی این حوزه را پوشش می‌دهند:  
• یادگیری گراف از داده‌ها:  
Dong و همکاران (2016) روش‌هایی برای یادگیری لامینای گراف بر اساس فرض نرمی سیگنال ارائه کردند. به طور مشابه، Kalofolias (2016) مدل‌هایی برای یادگیری گراف‌های کم‌حجم و پراکنده معرفی کرد.  
• پردازش سیگنال گرافی:  
Ortega و همکاران (2018) مروری جامع بر GSP داشتند، که بر چگونگی استفاده از ساختارهای گراف برای پردازش و تحلیل سیگنال‌هایی که در دامنه‌های نامنظم قرار دارند، تأکید می‌کرد.  
• ساختن گراف کلاسیک:  
گراف‌های k-NN، ε-نیرهاودگراف‌ها و گراف‌های کاملاً وصل با لبه‌های وزن‌دار، مدت‌ها برای تقریب زدن لباس‌پوش‌های داده (مثلاً در یادگیری منی‌فولد مانند Isomap و Laplacian Eigenmaps) استفاده شده‌اند.  
• استنتاج توپولوژی:  
در سال‌های اخیر، مدل‌های استنتاج توپولوژی (مثلاً Egilmez و همکاران، 2017) ظهور یافته‌اند که تمرکز بر استنتاج ساختار گراف از طریق بهینه‌سازی‌هایی دارند که ویژگی‌های داده‌های مشاهده‌شده را تطابق دهند.

اهداف تحقیق  
• هدف ۱: توسعه روش‌های سیستماتیک برای نقشه‌برداری داده‌های ورودی به ساختار گرافی بهینه، چه directed و چه undirected.  
• هدف ۲: بررسی کاربرد ابزارهای پردازش سیگنال گرافی (مانند تبدیل‌های فوریه گراف، فیلتر کردن) در گراف‌های ساخته‌شده.  
• هدف ۳: ارزیابی تأثیر روش‌های مختلف ساخت گراف بر وظایف پایین‌دست مانند خوشه‌بندی، پیش‌بینی و نویززدایی.

روش‌های پیشنهادی  
• تکنیک‌های ساخت گراف:

* روش‌های مبتنی بر فاصله (مثلاً k-NN، ε-گراف‌ها)
* یادگیری گراف با استفاده از بهینه‌سازی (مانند یادگیری لاپلاسین گراف)
* رویکردهای احتمالاتی و نظریه اطلاعات (مثل گراف‌های بر اساس هم‌اطلاعی mutual information)  
  • ابزارهای نظریه گراف:
* خصوصیات طیفی (مقدارهای ویژه لاپلاسین یا ماتریس‌های همسایگی)
* شاخص‌های مرکزیت و اتصال
* تحلیل ناپایداری و کاهش گراف  
  • ابزارهای پردازش سیگنال گرافی:
* معیارهای نرمی سیگنال در گراف‌های ساخته‌شده
* تحلیل تبدیل فوریه گرافی سیگنال‌های داده‌ها
* طراحی فیلترهای گراف برای نویززدایی و استخراج ویژگی‌ها

پیشنهادهای محتمل  
• چارچوبی واحد برای تبدیل داده‌ها به گراف برپایه مبانی نظری و خواص تجربی.  
• معیارهای نوین برای ارزیابی «کیفیت» گراف‌های ساخته شده جهت وظایف پردازش سیگنال.  
• کاربردهای جدید در حوزه‌هایی مانند شبکه‌های حسگری، داده‌های زیستی و تحلیل شبکه‌های اجتماعی.

چالش‌های مورد انتظار  
• تعادل بین ناپایداری گراف و اتصال آن‌ها.  
• حساسیت به نویز و خروجی‌های ناسازگار در داده‌ها.  
• مقیاس‌پذیری برای کار با مجموعه داده‌های بزرگ