

Thesis Description

EGOR DMITRIEV, Utrecht University, The Netherlands

1 DESCRIPTION

Recent developments in big data and graph representation learning have allowed researchers to make breakthroughs in social network analysis and the identification of communities. While opening a lot of research opportunities, such approaches are highly limited to snapshots of rapidly evolving social networks.

Relying on novel research of dynamic graph representation learning, the goal of my thesis project is to build a framework for community detection and representation in evolving heterogeneous networks. To verify the merit of the proposed framework, it will be evaluated against baselines on static heterogeneous graphs, and analyzed against gathered twitter dataset on covid measures (citation needed).

The challenges addressed in this project include but are not limited to the fact that the covered subjects are still novel and under active research. Because of that, there is still no clear consensus for definitions of some important structures such as temporal communities and dynamic networks. Over the course of the project, their definitions need to be thoroughly researched and compiled from related works and the to be solved problem. The nature of networks containing multimodal data in form of node types, edge types, and temporal differences also needs to be addressed and solved appropriately. Finally, the dynamicity of the communities needs to be explored as these communities may emerge, grow, split and fade over time.

2 OUTLINE

- **Goal:** is to build a framework for community detection and representation in heterogeneous dynamic networks.
- **Our specific case:** build a framework for community detection, representation and tracking over time in twitter network.
- **Use cases:**
 - Detected communities and their representation can be used in downstream ML tasks such as:
 - * Sentiment detection
 - * Classification / Multi-label Classification of certain communities
 - * (Temporal) Link prediction
 - Higher level use cases are:
 - * Tracking of users/community sentiment regarding to covid measures
 - * Tracking of users changing interests
 - * Detecting of emerging communities for marketing and advertisement
 - * Detection of extremism on social platforms
- **Technicalities:**
 - Definition: “**community**” a group of nodes where the intra-group connections are denser than the inter-group ones.
 - * In this case community consists of twitter users, tweets and hashtags
 - * Community is distribution covering the entity space (semantically representing interests)
 - * Community is dynamic over time and may evolve covering a different part of the entity space

- Define “interest group” / “subcommunity” / “topic”? (a narrow definition of community which doesn’t evolve over time)
- Evolving: Graph data increases but does not decrease (absence of event edges)
- Dynamic: Graph data increase and decreases (presence of dynamic edges)
- **Challenges:**
 - Communities may evolve over time:
 - * Covering more or fewer interests
 - Heterogenous dynamic networks may contain different kind of nodes
 - Dynamic networks may contain different kind of edges:
 - * Static edges (always valid)
 - * Event edges (valid for a range of timesteps)
 - * Dynamic edges (valid since a timestep)