Thesis Description

EGOR DMITRIEV, Utrecht Univeristy, The Netherlands

1 DESCRIPTION

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

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

2 OUTLINE

- **Goal**: is to build a framework for community detection and representation in heterogenous 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 Calssification 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 definiton of community which doesn't evolve over time)
    - Evolving: Graph data increases but does not decrease (absence of event edges)
    - Dynamic: Graph data increae 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 containt different kind of edges:
  - * Static edges (always valid)
  - * Event edges (valid for a range of timesteps)
  - * Dynamic edges (valid since a timestep)