Web Structure Mining Assignment

Li Xiaoli

Project

- Project 1: Community and Key Player Detection
- Project 2: Your Own Project?

Project 1: Community and Key Player Detection

- **Objectives**: detect communities and corresponding key players for each community
- **Techniques**: all techniques, covered in our lectures and implemented in the tools (welcome using new techniques)
- **Suggested Data** (but not limited to)
  - flights.dat and routes.dat: http://openflights.org/data.html; AIRPORTS_URL = https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat

Project 1: Community and Key Player Detection

- The Global Terrorism Database (GTD) https://www.start.umd.edu/gtd/
- Disease data http://www.nature.com/articles/ncomms5212
- Explore deaths and battles from this fantasy world https://www.kaggle.com/mylesoneill/game-of-thrones
- You can also use another available network data for your project, e.g. your own Facebook data
- The last choice: choose one data set from Stanford Large Network Dataset Collection https://snap.stanford.edu/data/

Project 1: Community and Key Player Detection

- **Result expectation**
  - You should show the knowledge and insights that you got from the data using various network measures and techniques
  - Why they are useful?
  - Good visualization
  - Highlights (novel algorithms/techniques)

Project 1: Community and Key Player Detection (Cont.)

- **Possible technical solution 1**
  - Detect the communities from the network that you have chosen
  - Within each community, detect key players in terms of various evaluation metrics
  - You can use the existing methods to detect the community and rank the nodes or propose *novel* methods (for community and key player detection) customized to your network data.

Project 1: Community and Key Player Detection (Cont.)

- **Possible technical solution 2**
  - Detect key players from the networks
  - Use the key players as the seeds to include more community members to form the communities (seeds/ key players are the center of the formed communities).
  - This need you to design an algorithm and do implementation for seed expansion, including other nodes that are highly interactive with seeds and other included nodes.

Project 1: Community and Key Player Detection (Cont.)

- **Possible technical solution 3**
  - Propose something new method to *concurrently* detect key players and communities
  - This will need your own idea and implementation

Project 2

- You can propose your own project related to the network analysis
- Good with practical applications
- Send me the description by email to **xlli@i2r.a-star.edu.sg** and I will let you know if it is suitable as your project

Tools

- You can use any tools or combinations for network analysis and visualization, including but not limited to
  - Gephi
  - R
  - Cytoscape
  - Pajek
  - NetMiner
  - SNAP
  - …..

Your Own Idea and Teamwork

Group Presentation

- **Materials**
  - Send your presentation slides, short reports (<=10 pages), well documented codes to xlli@i2r.a-star.edu.sg by **Nov 2 (PT class)** and **Nov 6 (FT class)**.
- **Presentation**
  - Group presentation date: **Nov 3 (PT class)** and **Nov 7 (FT class)**
  - All the group members must present part of the slides
  - 10 min presentation (if presentation time exceeds 2 more mins, your score will be deducted)

Group Presentation

- **PPT Slide can cover the following content**
  - **Introduction** (motivation and problem definition)
  - **Related works**
  - **The method**
  - **Experimental Study and Analysis** (settings, results, insights and comparisons)
  - **Summary of Project Achievements**
  - **Future Directions for further Improvements**
  - **Implementation** (well-commented source codes)

Evaluation Criteria

- Interestingness of your Insights
- Novelty of your solution
- Clarity of your presentation
- Technical depth of your solution
- Innovation of your problem

Thank You

Contact: xlli@i2r.a-star.edu.sg if you have questions