Project Assignment

Li Xiaoli

Web Analytics Project Assignment

- Project 1: Community and Key Player Detection

- Project 2: Your Own Project related to Web Analytics content. Drop me an email or talk to me whether your proposed project is relevant.

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
- Bus/training network (Singapore)
- Movie databases
- Paper citation databases
- You can also use *another available network data* for your project, e.g. your own Facebook data, that you believe you can get interesting insights
- Choose one data set from Stanford Large Network Dataset Collection https://snap.stanford.edu/data/ [not recommended]

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 in practice?
  - Good visualization
  - Highlights (novel algorithms/techniques if any)

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

- **Possible technical solution**
  - 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 key players (usually rank the nodes within a community and find most important few, e.g. using degree, cc, betweenness, pagerank, authority/hub score etc) or propose *novel* methods (for community and key player detection) customized to your network data.

Project 2

- You can propose your own project related to the Web Analytics analysis

- Good with practical applications

- Send me the description by email to **xlli@i2r.a-star.edu.sg** and **xlli@ntu.edu.sg** , then 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 (each tool has its own characteristic, including but not limited to
  - Gephi
  - R
  - Python
  - Cytoscape
  - Pajek
  - NetMiner
  - SNAP
  - …..

Your Own Idea and Teamwork

Group Presentation

- **Materials**
  - Send your 1) presentation slides, 2) Word or PDF project report, and 3) source codes to **lixl@i2r.a-star.edu.sg** and **xlli@ntu.edu.sg** by **Feb 3 2020.**

- **Presentation**
  - Classroom project group presentation: **Feb 5 2020**
  - All the group members must present part of the slides, unless you have valid reasons
  - Around 12 min presentation

Group Presentation

- **PPT Slide and project report can cover**
  - **Introduction** (motivation and problem definition)
  - **Related works** (if any)
  - **The methods** (explain the basic idea and provide some examples to illustrate how it works – benefit everyone)
  - **Experiments** (settings, results and comparisons)
  - **Conclusion and recommendation** (summary of project achievements, why knowledge and insights found are useful? possible future improvements, what are the benefits of doing this project?)
  - **Implementation** (well-commented source codes, explain a bit details to benefit everyone)

Project Evaluation Criteria

- Interestingness of your problem
- Sufficient insights and knowledge
- Clarity of your presentation
- Well-documented codes
- Novelty & technical depth of your solution

Thank You

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