Team 81 – Analyzing the Implicit Social Network from Github Activities

- Ran Tavory
- Ruzvidzo Ngulube
- Jonathan del Campo
- Brendan Danyluik (dropping)

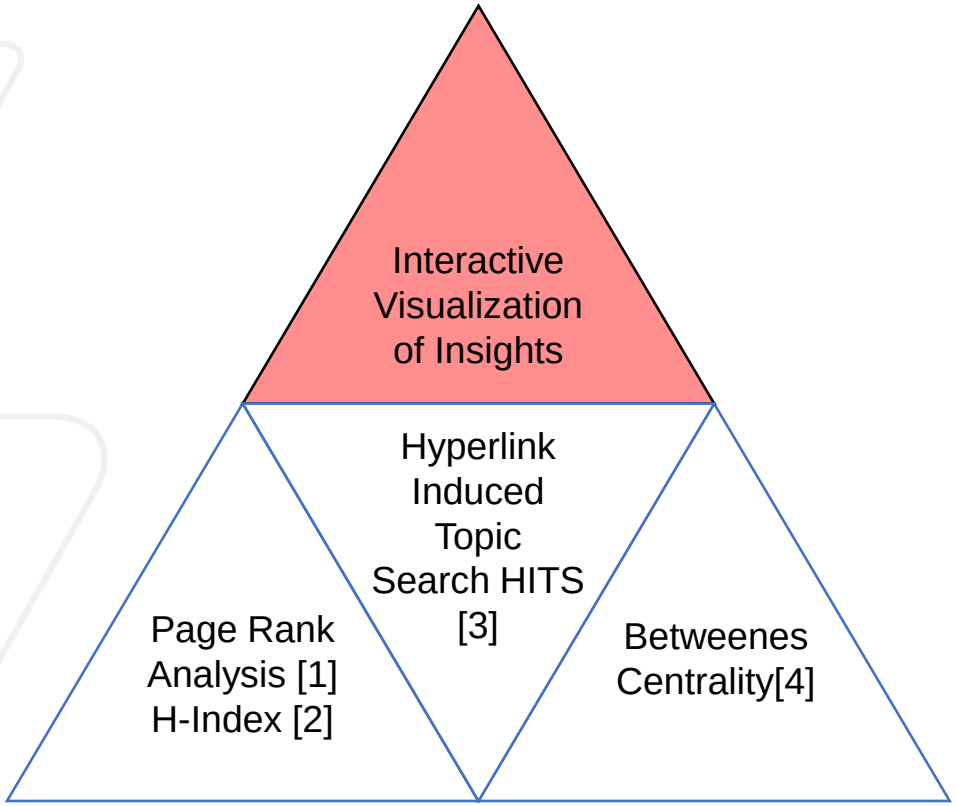
Date 2022-10-10

Georgia Tech

Proposal Presentation

How is it done today? Limits

Several **statistical procedures** to rank influence networks and users showed in **static graphs**



What are you trying to do?

Interactive graph UI that allows to find:

- **shortest paths** between users through project **co-activity**
- **most influential users** in a specific technology scope
- **rate users** based on social developer distance to influencers developers.

Based on **Github Data**

Georgia Tech

Proposal Presentation

What's new in the approach?

Why will it be successful?

- Interaction **accelerates** insights discovery.
- **Earlier detection** of emerging communities for business headhunters
- Better assessment of **user contribution** in Github community

Who cares?

- **Business headhunters** and marketing specialists to convince software developers influencers
- Users for **personal branding**
- **GH developers**

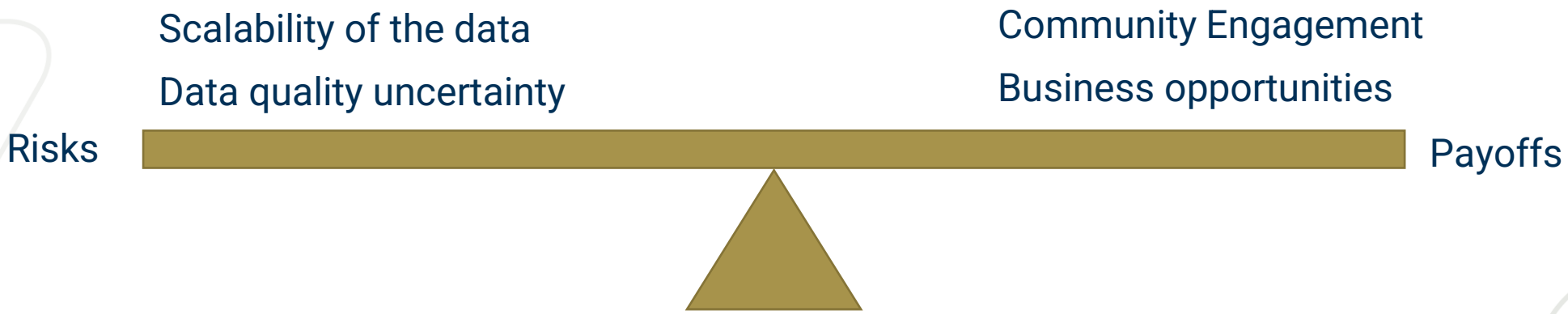
Georgia Tech

Proposal Presentation

Difference and impact. Impact measurement.

- Project UI **interaction enables** to attract more users to make the **tool popular** for influence network analysis purposes.
- The tool can be fundamental to **evidence social prerequisites** in big projects in software development [5]
- **Project mentions** in social networks as key to measure the **impact over the community**

Risks and payoffs



Georgia Tech

Proposal Presentation

How much will it cost?

Requirements:

- **Data storage** in a solution suited for graphs, like Neo4j
- **Data processing** on Python or Spark (feasibility analysis)
- **User Interface development**

On commercial mode, the solution could cost around $150- $ 1150/month approximately (quick estimation) depending of the scale, to serve database and UI web server

How long will it take?

Considering some parallelization of tasks:

- **Data collection:** 4 weeks
- **Data processing / analytics:** 3 weeks
- **UI development:** 4 weeks

Midterm and final check and progress measurement

Milestone set up and peer review to guarantee the successful implementation of the project.

Work item	Main responsible	Start	Duration
Raw Data Collection	Jonathan del Campo	November 1	4 weeks
Data Augmentation / Analytics	Ruzvidzo Ngulube	November 7	3 weeks
Web and UI	Ran Tavory	November 1	4 weeks

Georgia Tech

References

- [1] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999-66. Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/ Previous number = SIDLWP-1999-0120
- [2] Yan Hu, Shanshan Wang, Yizhi Ren, and Kim-Kwang Raymond Choo. 2018. User influence analysis for Github developer social networks. Expert Systems with Applications 108 (2018), 108–118. https://doi.org/10.1016/j.eswa.2018.05.002
- [3] Yan Hu, Jun Zhang, Xiaomei Bai, Shuo Yu, and Zhuo Yang. 2016. Influence analysis of Github repositories. https://doi. org/10.1186/s40064-016-2897-7
- [4] Linton C. Freeman. 1977. A Set of Measures of Centrality Based on Betweenness. Sociometry 40, 1 (1977), 35–41. http: //www.jstor.org/stable/3033543
- [5] Casey Casalnuovo, Bogdan Vasilescu, Premkumar Devanbu, and Vladimir Filkov. 2015. Developer Onboarding in GitHub: The Role of Prior Social Links and Language Experience. In Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering (Bergamo, Italy) (ESEC/FSE 2015). Association for Computing Machinery, New York, NY, USA, 817–828. https://doi.org/10.1145/2786805.2786854

Georgia Tech