Creating Socio-Technical Patches for Information Foraging: A Requirements Traceability Case Study

Darius Cepulis

Dr. Nan Niu, Chair
Dr. Carla Purdy
Dr. Chia Han

Thesis submitted to the Faculty of the University of Cincinnati in partial fulfillment of the requirements for the degree of Master of Science in Computer Engineering

University of CINCINNATI

1

Overview

1. Introduction
2. Background

3. Examining Requirements Socio-Technical Graphs
4. Creating Socio-Technical Patches for Information Foraging

5. Discussions
6. Future Work

University of CINCINNATI

2

INTRODUCTION:
TERMINOLOGY AND PROBLEM

University of CINCINNATI

3

~~Information~~ Optimal Foraging Theory

[Image] (cc) https://www.flickr.com/photos/pavdw/19456699639

University of CINCINNATI

4

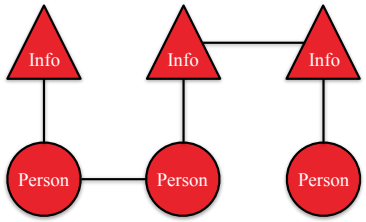
Examples of Information Patches

- Web Search: Web Site [1]
- Bug Fixing: Code Fragment [2]

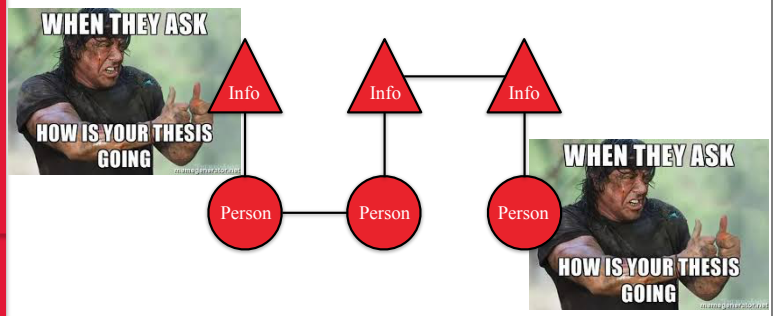
[1] E. H. Chi, P. Pirolli, K. Chen, and J. E. Pitkow, "Using information scent to model user information needs and actions and the web," in Conference on Human Factors in Computing Systems (CHI), Seattle, WA, USA, March-April 2001, pp. 490–497.
[2] J. Lawrance, C. Bogart, M. M. Burnett, R. K. E. Bellamy, K. Rector, and S. D. Fleming, "How programmers debug, revisited: an information foraging theory perspective," IEEE Transactions on Software Engineering, vol. 39, no. 2, pp. 197–215, February 2013.

Socio-Technical Systems

- Facebook
- YouTube
- Twitter
- GitHub
- Wikipedia



Socio-Technical Systems
What's a Patch?



Socio-Technical System:

Requirements Traceability

- Examining the life of a requirement from inception to implementation
- USDA Product Safety doubted [1]
- Facebook CEO can't explain decisions [2]
- Socio-Technical

[1] P. Mäder, P. L. Jones, Y. Zhang, and J. Cleland-Huang, "Strategic traceability for safety-critical projects," IEEE Software, vol. 30, no. 3, pp. 58–66, May/June 2013.
[2] Q. Forgey and A. E. Weaver, "Key moments from Mark Zucker- berg's senate testimony," https://www.politico.com/story/2018/04/10/ zuckerberg-senate-testimony-facebook-key-moments-512334?cid-apn, April 2018, accessed: Apr 2018.

Requirements *Traceability Questions*

- ...are an information need. Prey.
- What's the patch?

- *where should a user search to understand their requirements traceability question?*

9

Traceability Patches

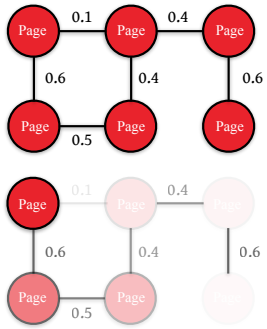
1. Examine 4 Requirements Repositories (with 125 Traceability Questions) as Socio-Technical Graphs
2. Identify what needs to be in a patch
3. Spreading Activation Algorithm to create patches as small as 5-10 nodes

10

BACKGROUND

Information Foraging Theory

- Pirolli & Card, Web User Flow by Information Scent (**WUFIS**) [1]
  - Nodes, Webpages (*patch*)
  - Edges, Links
  - Edges weighted by relatedness (*scent of other patches*)
  - Predicts where the user will/should navigate with spreading activation. (*which patch will I switch to?*)

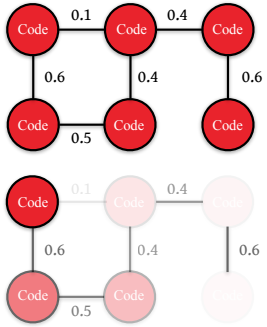


[1] P. Pirolli, "Computational models of information scent-following in a very large browsable text collection," in Conference on Human Factors in Computing Systems (CHI), Atlanta, GA, USA, March 1997, pp. 3–10.

12

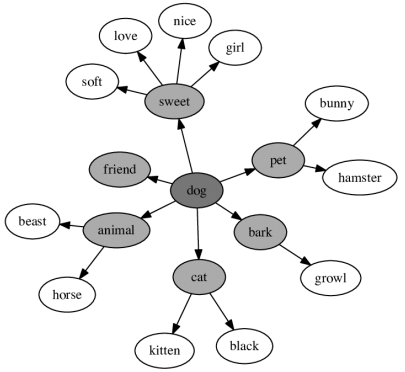
Information Foraging Theory

- **PFIS** [1]: WUFIS for Code Navigation
  - Nodes, Code (*patch*)
  - Edges, Links
  - Edges weighted by relatedness (*scent of other patches*)
  - Predicts where the user will/should navigate with spreading activation (*which patch will I switch to?*)



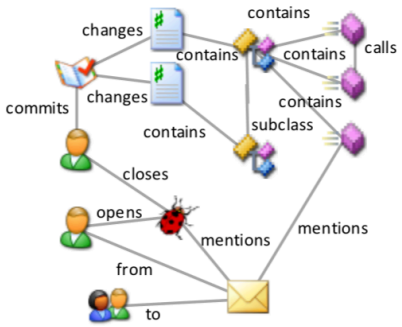
[1] J.Lawrance,R.K.E.Bellamy,M.M.Burnett,andK.Rector,"Usinginformationscent to model the dynamic foraging behavior of programmers in maintenance tasks," in *Conference on Human Factors in Computing Systems (CHI)*, Florence, Italy, April 2008, pp. 1323–1332.

Detour: Spreading Activation



Socio-Technical Topology

- Codebook [1]: People + Artifacts
- Regular Language Reachability



[1] JA. Begel and R. DeLine, "Codebook: social networking over code," in *International Conference on Software Engineering (ICSE)*, Vancouver, Canada, May 2009, pp. 263– 266.

- Can we Spread Activation over a
- Socio-Technical Graph like Codebook
- to make IFT Patches?

University of CINCINNATI

EXAMINING REQUIREMENTS SOCIO-TECHNICAL GRAPHS

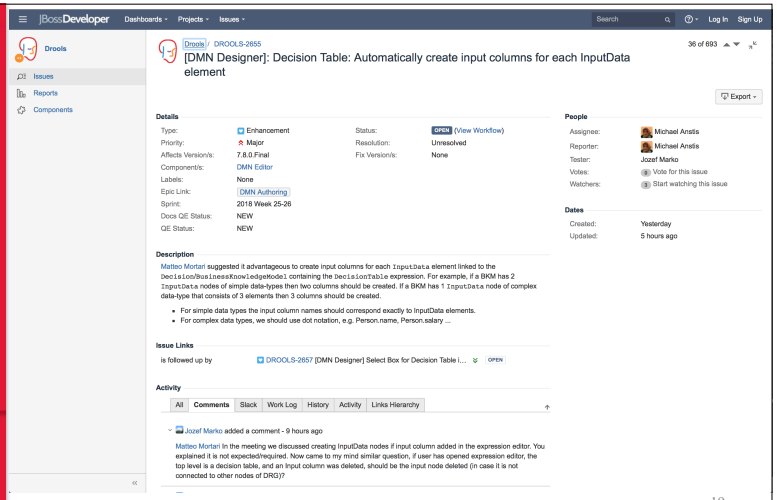
University of CINCINNATI

The Information Environment

- Where can we find Traceability Questions?
- Issue Trackers like Jira [1]

[1] P. Heck and A. Zaidman, "An analysis of requirements evolution in open source projects: recommendations for issue trackers," in *International Workshop on Principles of Software Evolution (IWPSE)*, Saint Petersburg, Russia, August 2013, pp. 43–52.

University of CINCINNATI



University of CINCINNATI

Traceability Questions in JIRA

- Identified Questions using CoreNLP [1]
- Manually identified answers

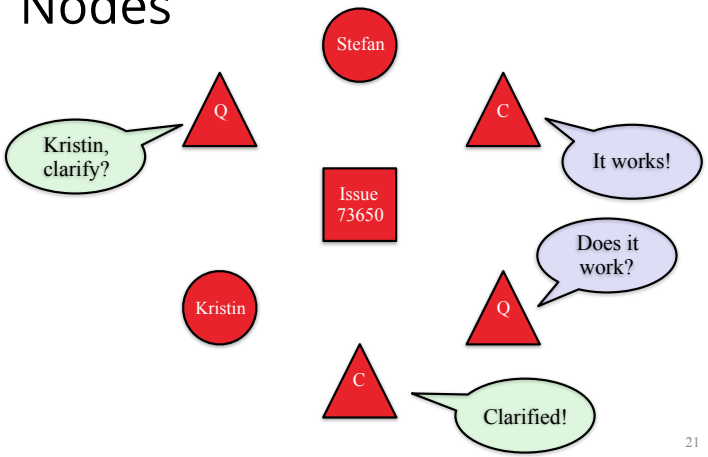
id	issue	body	asker	answered by	role	operations
1206	972	Can you please send a PR adding a new example...? …	Mario Fusco	Mauricio Salatino	Creator	add pull request
1220	963	This should be fixed for 6.4.0.Final, does it sound possible?	Mauricio Salatino	Petr Siroky	Assignee	change fix version
1371	907	...Please look at [my page]. What do you think…	Michael Kiefer	Geoffrey De Smet	Creator	marked as done
1377	905	Thanks...[~martenscs]. Is there any other prefix? …	Petr Siroky			
1383	901	any news here [~mfusco]?	David naranjo	Geoffrey De Smet		provide opinion
1627	823	Hi Geoff, I don't know what shoud be the expected behaviour? …	Michael Kiefer			

[1] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," in Association for Computational Linguistics (ACL) System Demonstrations, 2014, pp. 55–60. [Online]. Available: http://www.aclweb.org/anthology/P/P14/P14-5010

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Nodes



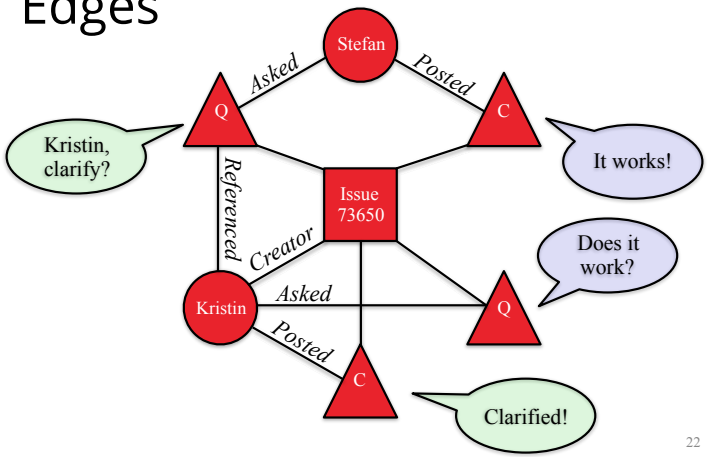
21

21

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Edges



22

22

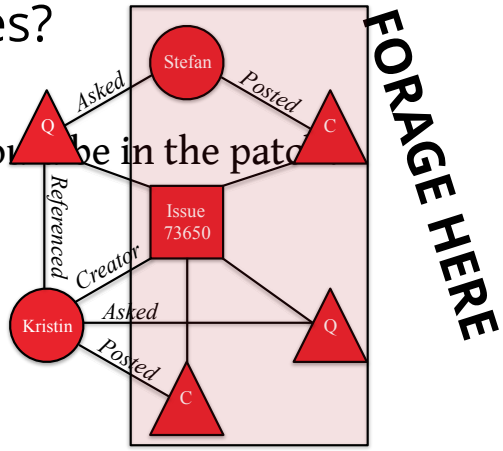
I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Patches?

No.

What sho[illegible]be in the patc[illegible]



FORAGE HERE

23

23

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Q&A Relationships

- *What should be in the patch?*
  A patch should probably include the answer
- Question: Need from a Patch
- Answer: Need may be satisfied
- *Where is the Answer?*

24

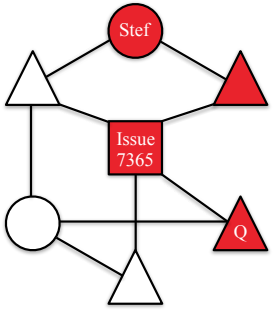
24

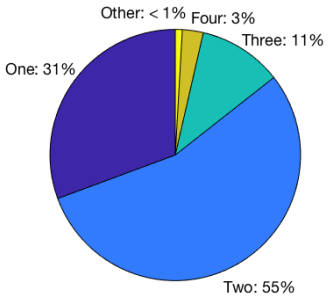
I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Q&A Relationships

- *Where is the Answer?*



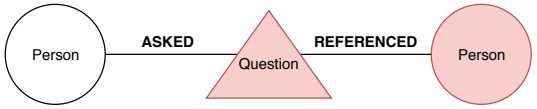


I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

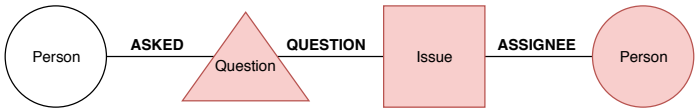
University of CINCINNATI

Where is the Answer? *One or Two Degrees*

1 Degree: Referenced or Self



2 Degrees: Creator or Assignee

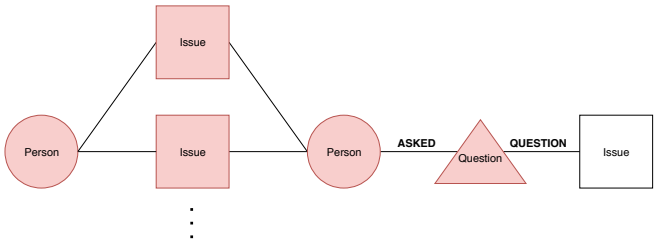


I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Where is the Answer? *Three or Four Degrees*

3 Degrees: Frequent Collaborator

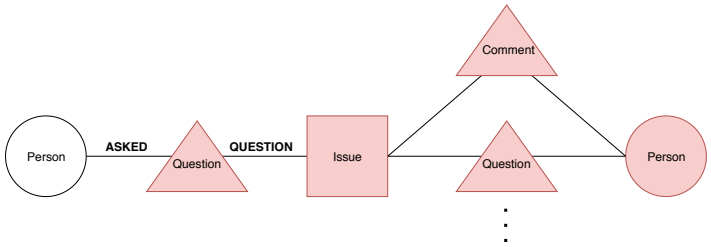


I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Where is the Answer? *Three or Four Degrees*

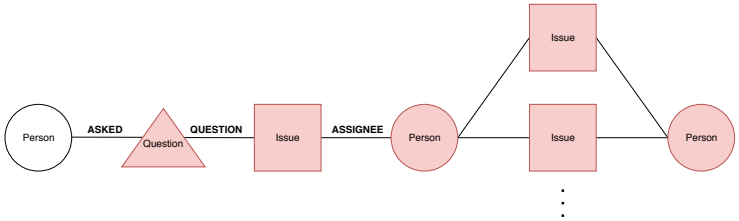
3 Degrees: Frequent Contributor



I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

Where is the Answer? *Three or Four Degrees*

4 Degrees: Frequent Collaborator of Creator or Assignee



University of CINCINNATI

29

29

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

Where is the Answer? *Five or More Degrees*

- More Frequent Collaborators or Contributors?

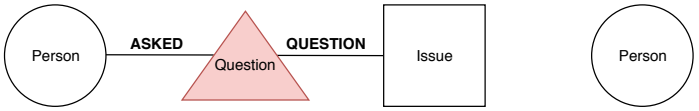
University of CINCINNATI

30

30

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

Where is the Answer? *Unconnected (For Now)*



University of CINCINNATI

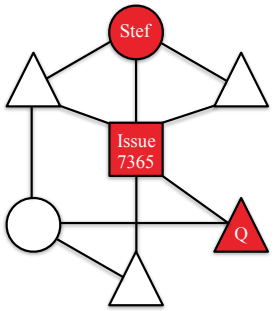
31

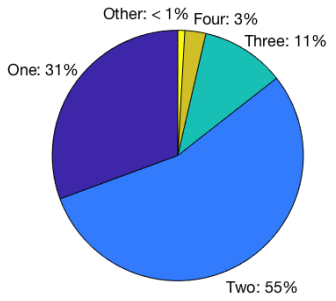
31

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

Q&A Relationships

- *A patch should probably include the answer*





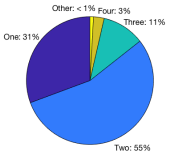
University of CINCINNATI

32

32

Q&A Relationships

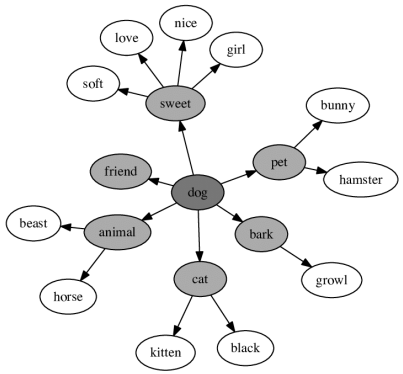
- *A patch should probably include the answer*



	Min	Q1	Med	Q3	Max
2 Degrees	4	29	192	442	1908
3 Degrees	5	271	510	1405	4192
4 Degrees	5	625	2288	3636	7542
5 Degrees	5	1949	3084	4597	7950

CREATING SOCIO-TECHNICAL PATCHES FOR INFORMATION FORAGING

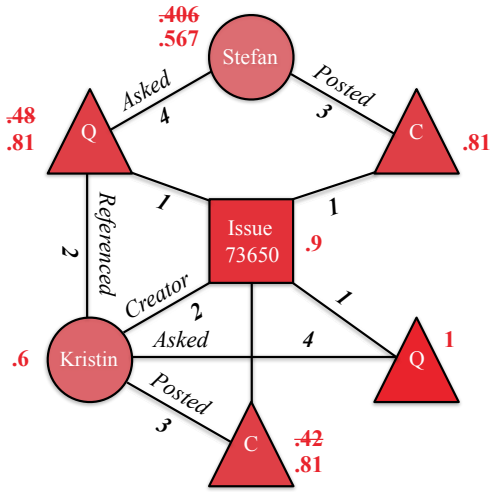
Spreading Activation



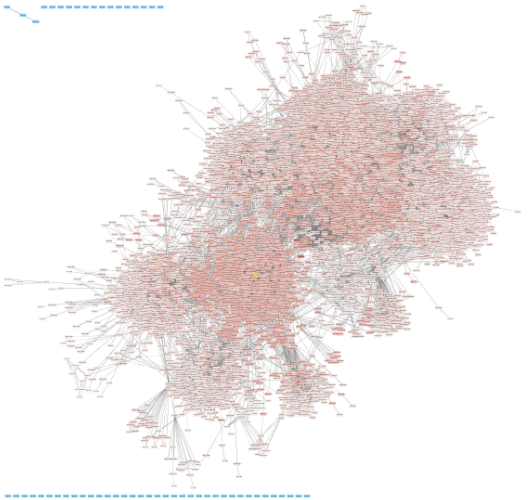
Spreading Activation: Weight

- Weight is Relatedness
- Weight is Relationships & Knowledge

- Comment to Issue: *Strong*
- Creator/Assignee to Issue: *Medium-Strong*
- Comment to Referenced: *Medium-Strong*
- Comment to User: *Medium*
- Question to User: *Weak*

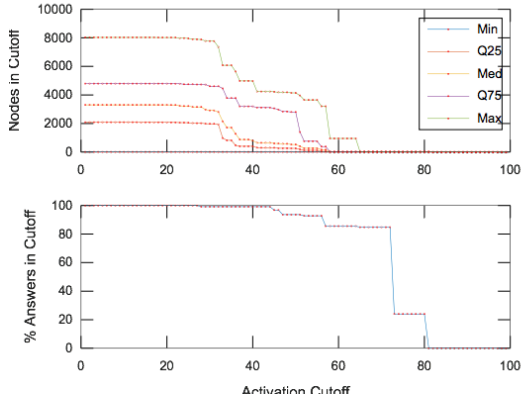


37



38

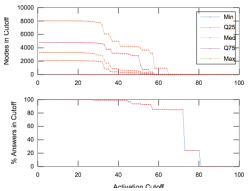
Delineating Patches: Activation



39

Delineating Patches: Activation

	Min	Q1	Med	Q3	Max	Answer
4 Degrees	5	625	2288	3636	7542	100%
$A \geq 0.45$	5	273	649	1860	4834	100%
$A \geq 0.50$	5	190	406	1399	3961	93%
$A \geq 0.56$	4	44	165	384	3207	85%
$A \geq 0.72$	2	4	6	10	24	84%



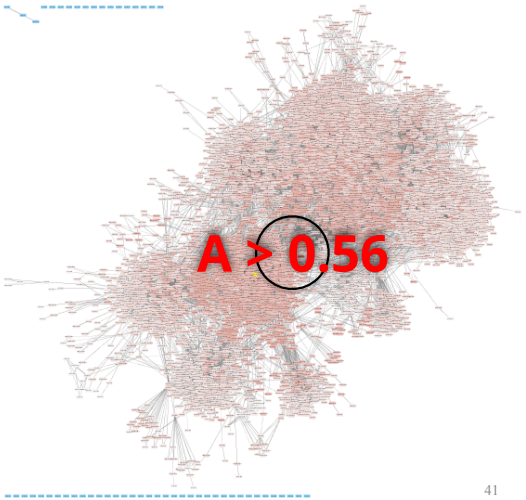
- All Answers at A > 0.45
- Frequent Collaborators at A > 0.50
- Frequent Contributors at A > 0.56
- Most Answers A > 0.72

40

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Practical Example



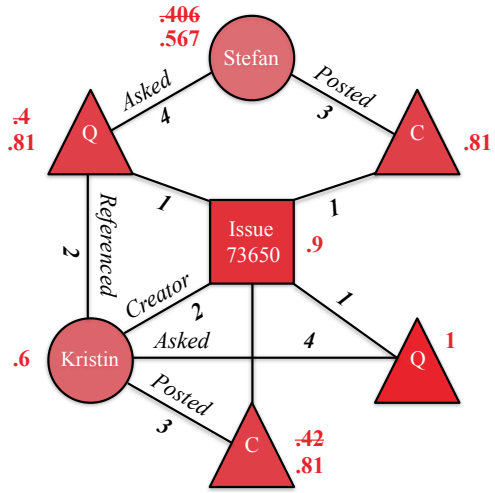
41

41

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Practical Example



42

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

DISCUSSIONS

43

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Implications

- Piorkowski et al. [1], Prey in Pieces
- RSTG Small World [2], 4 Degrees
- Answering Natural Language Traceability Needs [3]

[1] D. Piorkowski, A. Z. Henley, T. Nabi, S. D. Fleming, C. Scaffidi, and M. M. Bur- nett, "Foraging and navigations, fundamentally: developers' predictions of value and cost," in International Symposium on Foundations of Software Engineering (FSE), Seattle, WA, USA, November 2016, pp. 97–108.
[2] D. Chakrabarti and C. Faloutsos, "Graph mining: laws, generators, and algorithms," ACM Computing Surveys, vol. 38, no. 1, pp. Article 2, March 2006.
[3] P. Pruski, S. Lohar, W. Goss, A. Rasin, and J. Cleland-Huang, "TiQi: answering un- structured natural language trace queries," *Requirements Engineering*, vol. 20, no. 3, pp. 215–232, September 2015.

44

44

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

Conclusion

- Foraging Needs Patches
- We Made Patches
  - Identified Relationships
  - Spreading Activation
- This is a Foundation

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

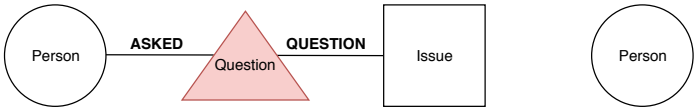
University of CINCINNATI

FUTURE WORK

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

University of CINCINNATI

New Data Types



- Connect Unconnected Users
- Shorter paths = tighter patches

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

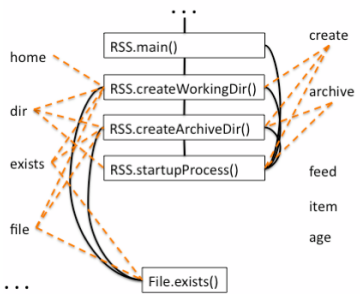
University of CINCINNATI

New Data Types

- Code & Commits
- Edges between Questions & Answers
- Semantic Similarity

New Data Types

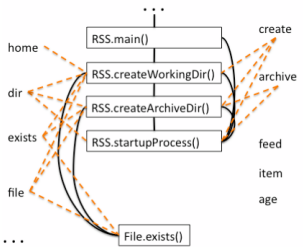
- Semantic Similarity



[Image] J. Lawrance, M. M. Burnett, R. K. E. Bellamy, C. Bogart, and C. Swart, "Reactive in- formation foraging for evolving goals," in *Conference on Human Factors in Computing Systems (CHI)*, Atlanta, GA, USA, April 2010, pp. 25–34.

New Data Types

- Semantic Similarity
- No shorter paths*
- No higher activation



* Comment 9732 –Issue 4548 – Person 4 – Issue 4792 – Person 691
Comment 9732 – **"workaround"** – Comment – Person 691

* Comment 1940 – Issue 1421 – Person 196 – Issue 1432 – Person 70
Comment 1940 – **"issue"** – Comment 2133 – Person 70

Better Algorithm

- Trial-Error Parameters
  - Decay
  - Weights
  - Frequency Reward
  - Patch Cutoff

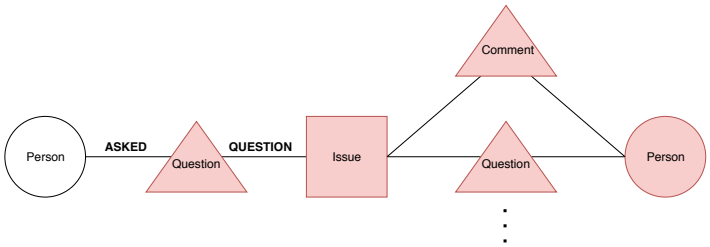
Better Algorithm

- Weights
- Frequency Reward
- Weightless
  - Answers in A > 0.72 Dropped 60%
- Weightless + Issue Activated
  - Similar to Original!

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

Better Algorithm

- Weights
- Frequency Reward



University of CINCINNATI

53

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

New Domains

- Requirements Traceability Case Study
- Applying Design Thinking…
  - Movies?
  - Safety Certification?
  - Composite Networks?

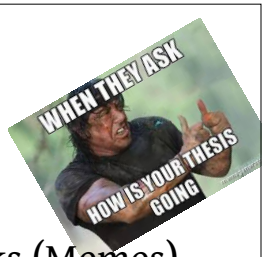
University of CINCINNATI

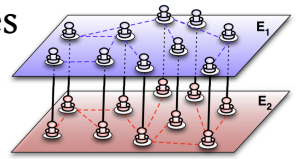
54

I. Introduction
II. Background
III. Examining Graphs
IV. Constructing Patches
V. Discussions
VI. Future Work

New Domains

- Composite Networks
  - Multiple JIRA Projects
  - Multiple Social Networks (Memes)
  - Popping News Bubbles

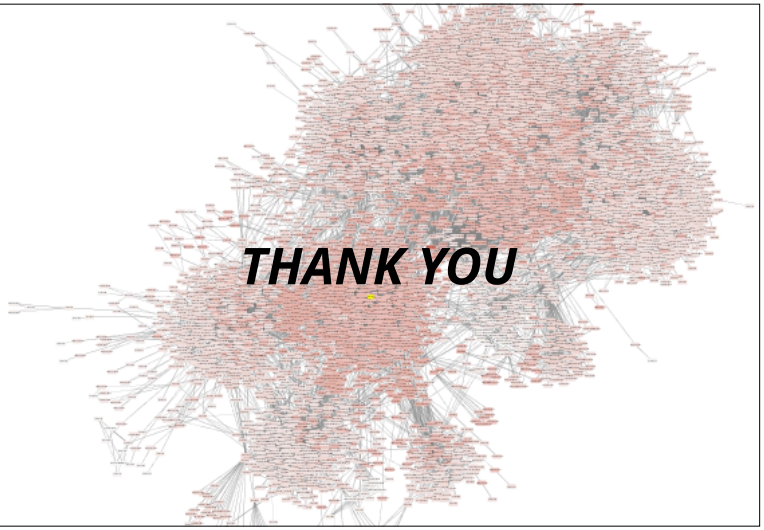




[Image] X. Wei, N. Valler, B. A. Prakash, I. Neamtiu, M. Faloutsos, and C. Faloutsos, "Com- peting memes propagation on networks: a case study of composite networks," ACM SIGCOMM Computer Communication Review, vol. 42, no. 5, pp. 5–12, 2012.

University of CINCINNATI

55



University of CINCINNATI