



proshanta sarkar <proshant.cse@gmail.com>

After XP2014

Subhajit Datta <subhajit.datta@gmail.com>
To: proshanta sarkar <proshant.cse@gmail.com>

Sun, Jul 20, 2014 at 12:09 PM

Dear Proshanta,

Great!

Could you now generate a developer-master-data for these 1040 developers (persons who own at least one review and are vertices of the latest DRCN)? Use the latest DRCN for the network related metrics.

1. Person id
2. Name
3. Email
4. IsBotAccount
5. Number of reviews owned by this person
6. Number of reviews owned by this person that are in status = merged
7. Number of reviews owned by this person that are in status = abandoned
8. Number of reviews commented upon by this person
9. Number of comments by this person across all reviews
10. Average number of comments per review by this person
11. Number of approvals by this person
12. Average approval score by this person (total approval scores divided by number of approvals)
13. Median approval score by this person
14. Time stamp of first comment
15. Time stamp of last comment
16. Comment span (difference in hours between first and last comment)
17. Time stamp of first approval
18. Time stamp of last approval
19. Approval span (difference in hours between first and last approval)
20. Degree of this person in DRCN
21. Betweenness of this person in DRCN
22. Clustering coefficient of this person in DRCN (Use function `transitivity()` in `igraph`)
23. Closeness centrality of this person in DRCN (Use function `closeness()` in `igraph`)
24. Eigenvector centrality of this person in DRCN (Use function `evcent()` in `igraph`)
25. Pagerank of this person in DRCN (Use function `page.rank()` in `igraph`)

Also, please use the following folder structure when you sync in GitHub

- networks (put all Pajek files here)
- reports (put all CSV files you generate here)
- code (put all code files here)
- info (put all other background information files here)

For the code, organize it like under subfolders like

- code
 - DRCN generator
 - Developer report generator
 - Review report generator

As you may have noticed, GitHub is a very powerful version control system. It is able to detect changes to the same file across all commits. If we keep all our materials well organized under GitHub, it will be easy to replicate our work in the future.

Let me know if you have any questions.

Regards,

Subhajit

—

Subhajit Datta, Ph.D.
www.dattas.net

Regards,

Subhajit

—

Subhajit Datta, Ph.D.
www.dattas.net

On Sun, Jul 20, 2014 at 1:23 PM, proshanta sarkar

[Quoted text hidden]

> ...

>

> [Message clipped]