

Some Thoughts on current Project: Highlight Key Nodes on Graph Layout

Liwenhan Xie, 2017.12.09

- I. PROPERTIES (suppose that keys are known/tagged)
 - A. Locally Hierarchy Structure:
Key \rightarrow Direct Neighbor \rightarrow Second Direct Neighbor $\rightarrow \dots$
 - B. Globally Complex Relationship:
bridge is of second priority yet they crossover; however the link relationship could be rather complicated when data is larger
- II. CLASSIFICATION of $G(V,E)$, where Point Of Interest $\subset V$
 - A. $|POI|$
 - 1. 10+: moderate pos (CURRENT FOCUS)
 - 2. 100+: clear group msg
 - 3. 1000+: impossible manual tagging, automatic generating POI is preferred; require cluster info and some interaction means
 - B. $|E|$
 - 1. sparse: ignore irrelevant nodes
 - 2. dense: bundle
 - C. $|V|$
 - 1. 1k+ (CURRENT FOCUS)
 - 2. 100k+
 - 3. 1000k+
- III. MAYBE RELEVANT RESEARCHES
 - A. Multilevel Algorithm – coarse the graph into the maximum independent vertex set or minimum dominant vertex set then lay the coarsest graph and do refinement steps through force directed algorithm [\[Paper\]](#)
 - B. 360.corp Sky-Eye – building complex network based on probability and statics theory
 - C. Large Graph Visualization Systems – cluster + interaction + multi-view... widely adopted $O(n)$ algorithms [\[Survey grouping\]](#) [\[Survey web\]](#)
 - D. ML – large graph construction approach to efficiently exploit all data points [\[Paper\]](#)
- IV. QUESTIONS

- A. I understand that we want to provide a nice layout with details, but what's the ultimate target? Give a delicate layout algorithm (I cannot imagine how is it possible to get satisfying results with only one algorithm due to various data), provide a comprehensive layout system (then there are so many things to concern, like useful interactive functions, explicit ways processing data in various form...), or just to set up a raw layout algorithm based on path info to show all the link-node information (what's the point of displaying all thousand nodes...why not shrink those unimportant nodes ?)
- B. Is the classification above too elaborate? But I don't assume that they are the same. And it seems to me that papers about large graph layout algorithms and complex network are talking different languages.
- C. Currently I am using the force directed algorithm to DBLP data, however as mentioned in some papers, this only ensure local optima, which is not what we want. And laying POI & bridges initially helps little (adjusting parameter may work but it can't solve the real problem... ML to adjust?)... I am having no idea what to do now.
- D. Here the key nodes are entitled supreme, then is there need to consider nodes tagged second interested?

V. APPLICATION (just fancy)

- A. Drug Manufacture: find out essential ingredient based on independent tests
- B. Academy Track: scholar collaboration, topic progression
- C. Security: Tracking suspects
- D. Visualization of Knowledge Graph, okay maybe ABC are only specific applications of this one.