

Some Graph Data

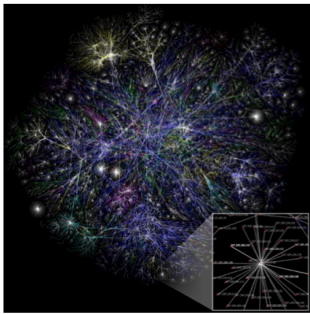


FIGURE: Graphical representation of webpage linkage

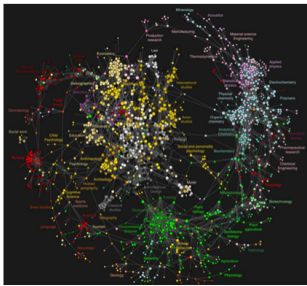


FIGURE: Graphical representation of relationships of scientific journals

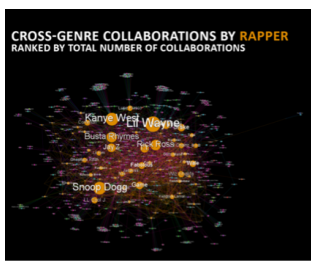


FIGURE: Network of collaborations among rappers



FIGURE: Network of US airlines

Biological Graph Data

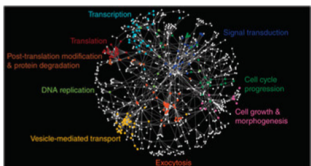


FIGURE: Gene Regulatory Network (GRN). The mRNA concentration follows a dynamic process (e.g. ODE) controlled by other related genes.

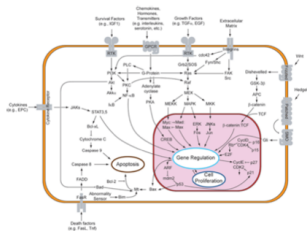


FIGURE: Cell Signal Transcriptional Network

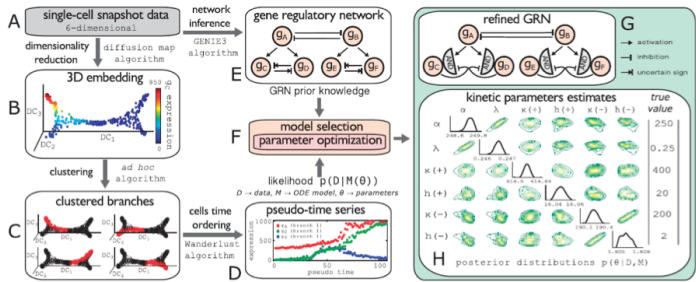


FIGURE: Framework of GRN produced from single-cell data.

Linking Websites

- Scoring websites by counting number of links
- Rescoring (reweighting) by considering the importance of the websites



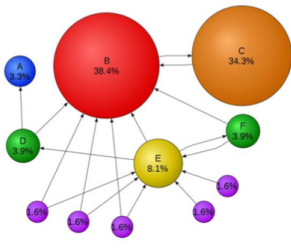
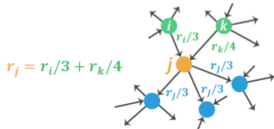
What sites link to pku.edu.cn?	What sites link to tsinghua.edu.cn?																								
Total Sites Linking In: 6,649	Total Sites Linking In: 8,579																								
<table border="1"> <thead> <tr> <th>Site</th><th>Page</th></tr> </thead> <tbody> <tr><td>1. baidu.com</td><td>bai.baidu.com/publication.html</td></tr> <tr><td>2. msn.com</td><td>msn.com/de-yehachrichte/news/number...</td></tr> <tr><td>3. qq.com</td><td>edu.qq.com/school</td></tr> <tr><td>4. hupu.com</td><td>bbs.hupu.com/4788328.html</td></tr> <tr><td>5. 163.com</td><td>biz.163.com</td></tr> </tbody> </table>	Site	Page	1. baidu.com	bai.baidu.com/publication.html	2. msn.com	msn.com/de-yehachrichte/news/number...	3. qq.com	edu.qq.com/school	4. hupu.com	bbs.hupu.com/4788328.html	5. 163.com	biz.163.com	<table border="1"> <thead> <tr> <th>Site</th><th>Page</th></tr> </thead> <tbody> <tr><td>1. yahoo.com</td><td>travelinspirations.yahoo.com/post?id...</td></tr> <tr><td>2. baidu.com</td><td>tieba.baidu.com/p/1e-vf-68&ie=//清华大学</td></tr> <tr><td>3. msn.com</td><td>msn.com/en-us/travel/tips/asia-the-bea...</td></tr> <tr><td>4. yandex.ru</td><td>ftp.yandex.ru/debian/README.minors.M...</td></tr> <tr><td>5. qq.com</td><td>city.qq.com</td></tr> </tbody> </table>	Site	Page	1. yahoo.com	travelinspirations.yahoo.com/post?id...	2. baidu.com	tieba.baidu.com/p/1e-vf-68&ie=//清华大学	3. msn.com	msn.com/en-us/travel/tips/asia-the-bea...	4. yandex.ru	ftp.yandex.ru/debian/README.minors.M...	5. qq.com	city.qq.com
Site	Page																								
1. baidu.com	bai.baidu.com/publication.html																								
2. msn.com	msn.com/de-yehachrichte/news/number...																								
3. qq.com	edu.qq.com/school																								
4. hupu.com	bbs.hupu.com/4788328.html																								
5. 163.com	biz.163.com																								
Site	Page																								
1. yahoo.com	travelinspirations.yahoo.com/post?id...																								
2. baidu.com	tieba.baidu.com/p/1e-vf-68&ie=//清华大学																								
3. msn.com	msn.com/en-us/travel/tips/asia-the-bea...																								
4. yandex.ru	ftp.yandex.ru/debian/README.minors.M...																								
5. qq.com	city.qq.com																								

Scoring the Pages

Ranking the webpage j by computing r_j :

$$r_j = \sum_{i \rightarrow j} \frac{r_i}{d_i}, \text{ (stream equation)}$$

where d_i is the out-degree of freedom of node i



Eigenvalue Problem

- $\mathbf{M}\mathbf{r} = \mathbf{r}$
- Use power method to solve for $\mathbf{r} = \lim_{t \rightarrow \infty} \mathbf{r}^{(t)}$:
 1. Initialization : $\mathbf{r}^{(0)} = (\frac{1}{N}, \dots, \frac{1}{N})^T$
 2. Iteration : $\mathbf{r}^{(t+1)} = \mathbf{M}\mathbf{r}^{(t)}$
 3. Stopping rule : $\|\mathbf{r}^{(t+1)} - \mathbf{r}^{(t)}\| \leq \epsilon$.
- Random walk interpretation : $\mathbf{r}^{(t)} = (r_i^{(t)})_i$ is a probability distribution, where $r_i^{(t)}$ represents the probability that the explorer stays in the webpage i at time t ; he randomly choose the next webpage according to the probability indicated by the matrix \mathbf{M} .
- This produces a Markov chain. And \mathbf{r} is its stationary distribution if \mathbf{M} is irreducible and non-periodic by Perron-Frobenius theory.

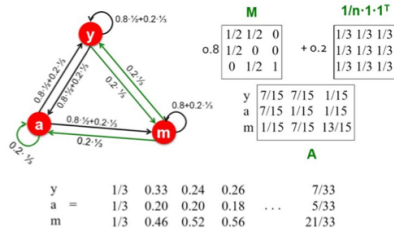
\mathbf{M} 是马尔可夫矩阵

Google PageRank

- To avoid spider traps (out-link absorbed by a small subset) and dead ends (no out-link), Google introduced the random page transition (Brin-Page,98) :

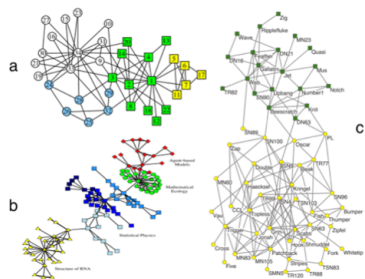
$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

- $\mathbf{A} = \beta \mathbf{M} + (1 - \beta) \frac{1}{N} \mathbf{1}\mathbf{1}^T$ is irreducible and non-periodic



Community Detection

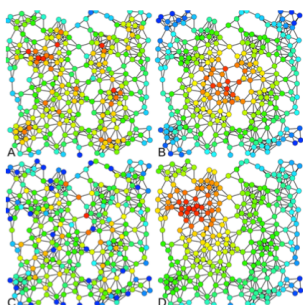
- Club organization from individuals (karate club)
- Collaboration network
- Social behavior of zebra



Centrality (Geometry of the Graph)

- Degree (or normalized by the total number of vertices) centrality : the number of edges linking the node
- Farness and closeness centrality, harmonic Centrality
- Betweenness centrality : the number of shortest paths passing through the current node
- Eigenvector centrality : \mathbf{r} in PageRank

按图中结点的度分



Community Detection Algorithm

- Hierarchical clustering based algorithms :
 - Girvan-Newman Algorithm
 - Improved by Newman's fast algorithm : A concept of "modularity" Q is introduced, agglomerate the subsets by maximizing ΔQ
- Fast Unfolding by V. D. Blondel, implemented in Gephi

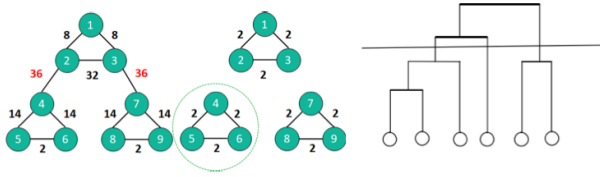


FIGURE: Left : Girvan-Newman ; Right : Newman's fast algorithm