复杂网络间节点匹配算法研究

报告人：宣琦
导师：吴铁军教授
Email：crestxq@hotmail.com
浙江大学 控制科学与工程学系

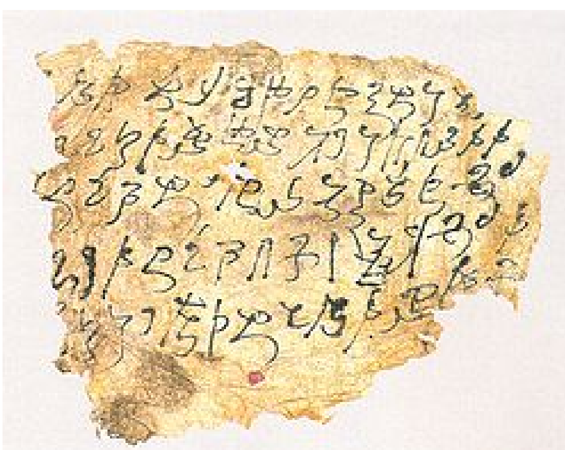
应用背景——几个例子

- 语言网络节点匹配
- 蛋白质关联网络节点匹配
- 社会网络节点匹配 ★

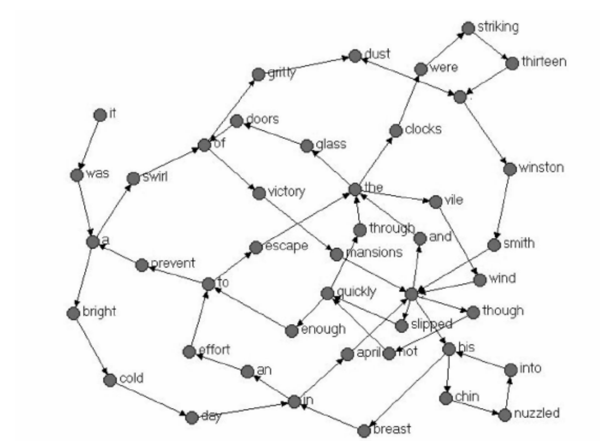
语言网络——古文字的破译

古埃及文

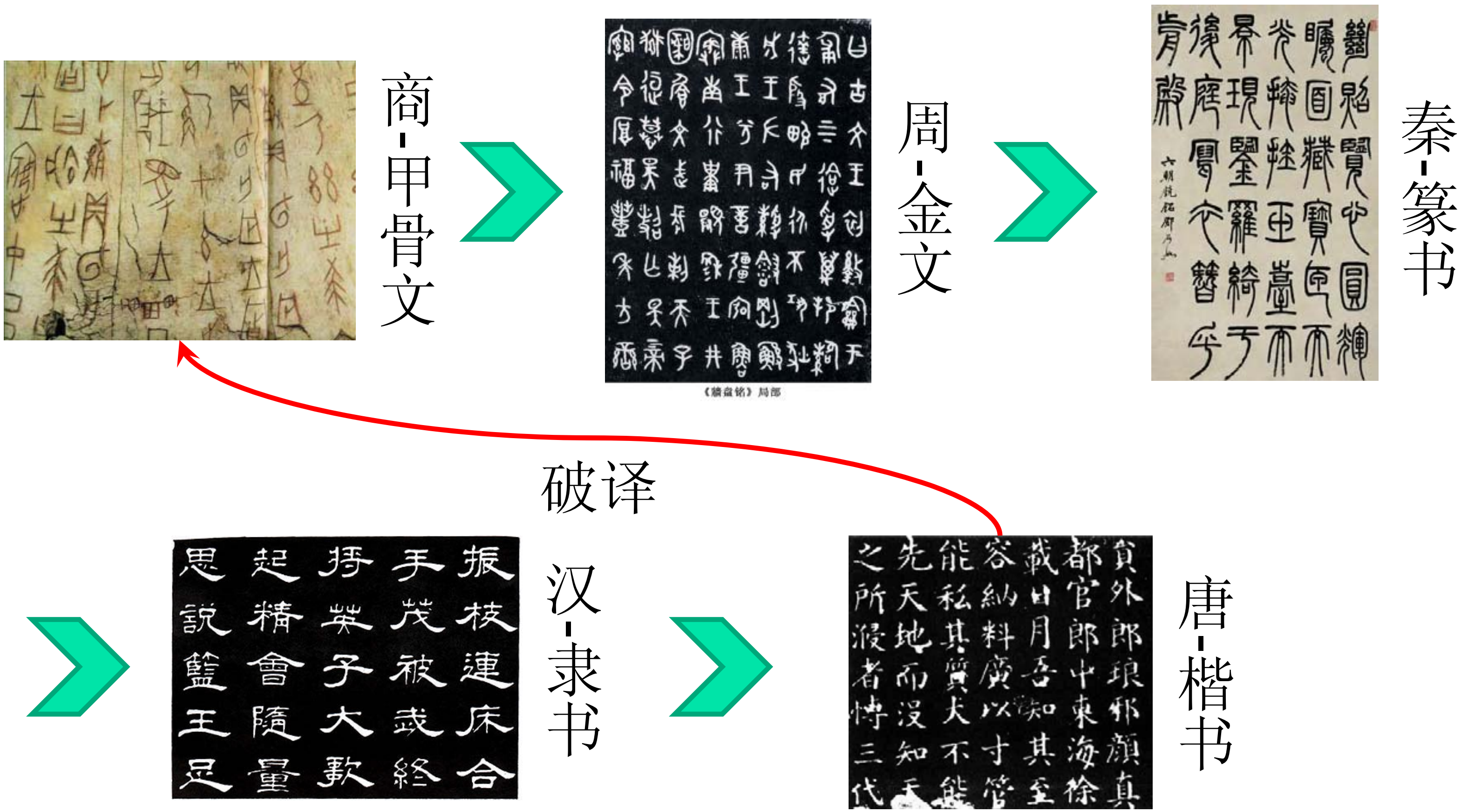
古巴比伦文



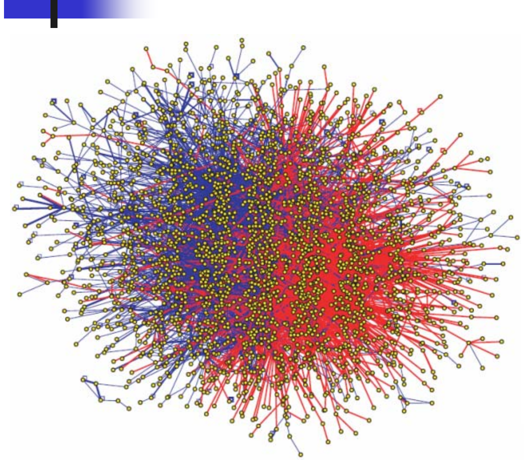
古印度文



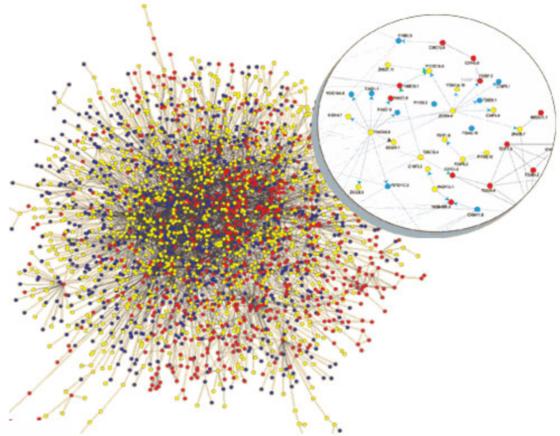
语言网络——古文字的破译



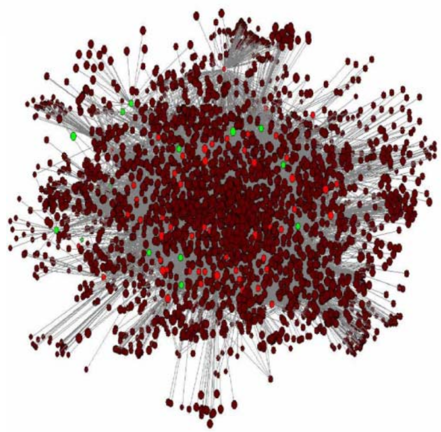
生物网络——同源蛋白质匹配



Human



C.Elegans

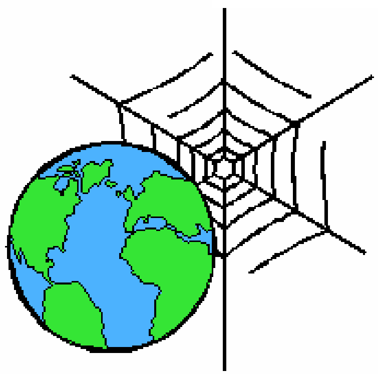


S.Cerevisiae

Yeast

匿名社会网络



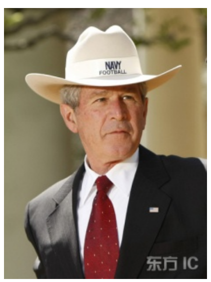



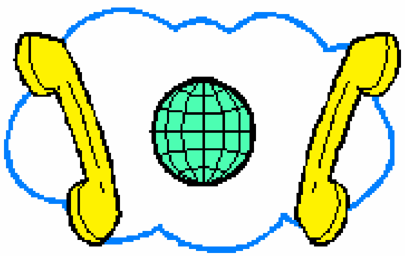


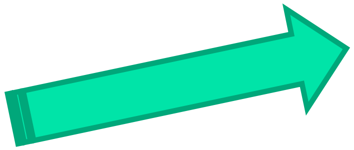

匿名网络：
MSN
QQ
Email
Twitter
Facebook
Skype
魔兽世界
…
…
…

大都不涉及利益关系

实名社会网络





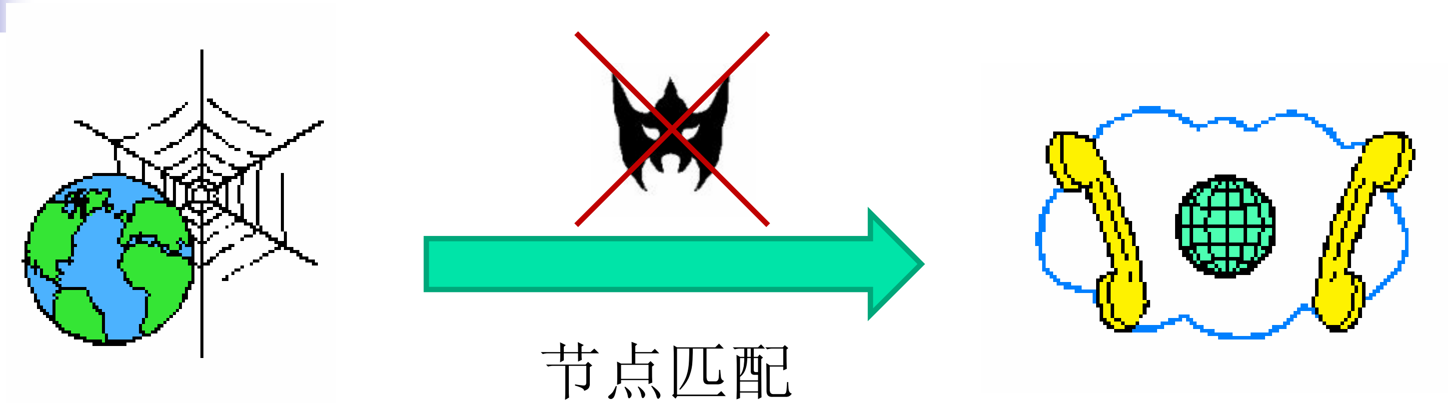


实名网络：

电话网络
银行账户网络
合著网络
电子商务
…
…
…

涉及利益关系

社会网络——个体身份匹配



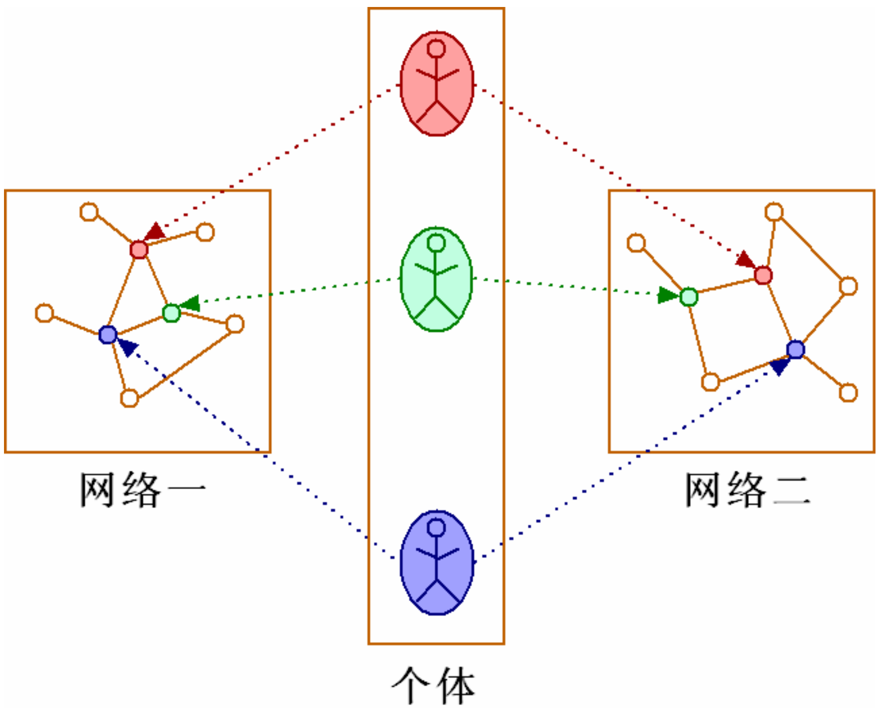
网络社会管理：法制 VS 人治

匿名网络设置实名信息 → 形成已匹配节点对（网络传感器）

→ 揭示其他人士在匿名网络和实名网络之间的身份匹配

匹配算法设计 ★

实行节点匹配的基本前提



生物网络中的同源蛋白质
语言网络中的语义
社会网络中的人
个体：

个体在不同网络中具有类似的连接行为

语言网络和生物网络：缓慢变异

一个社会网络例子

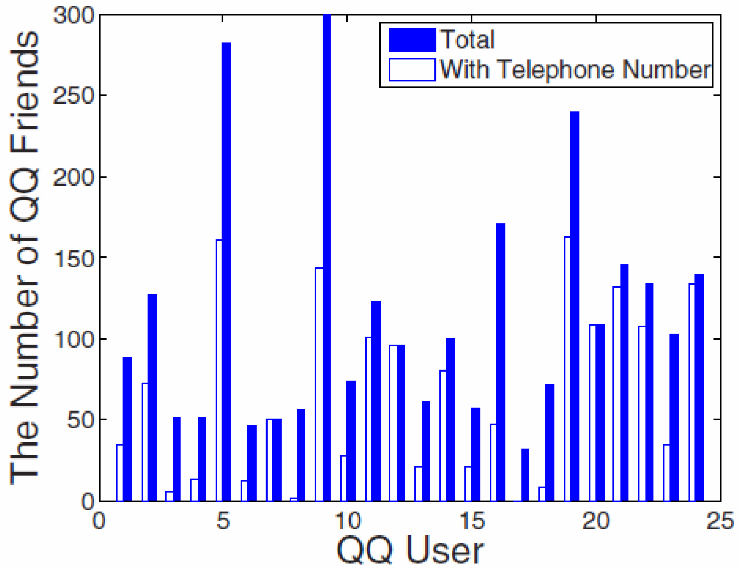
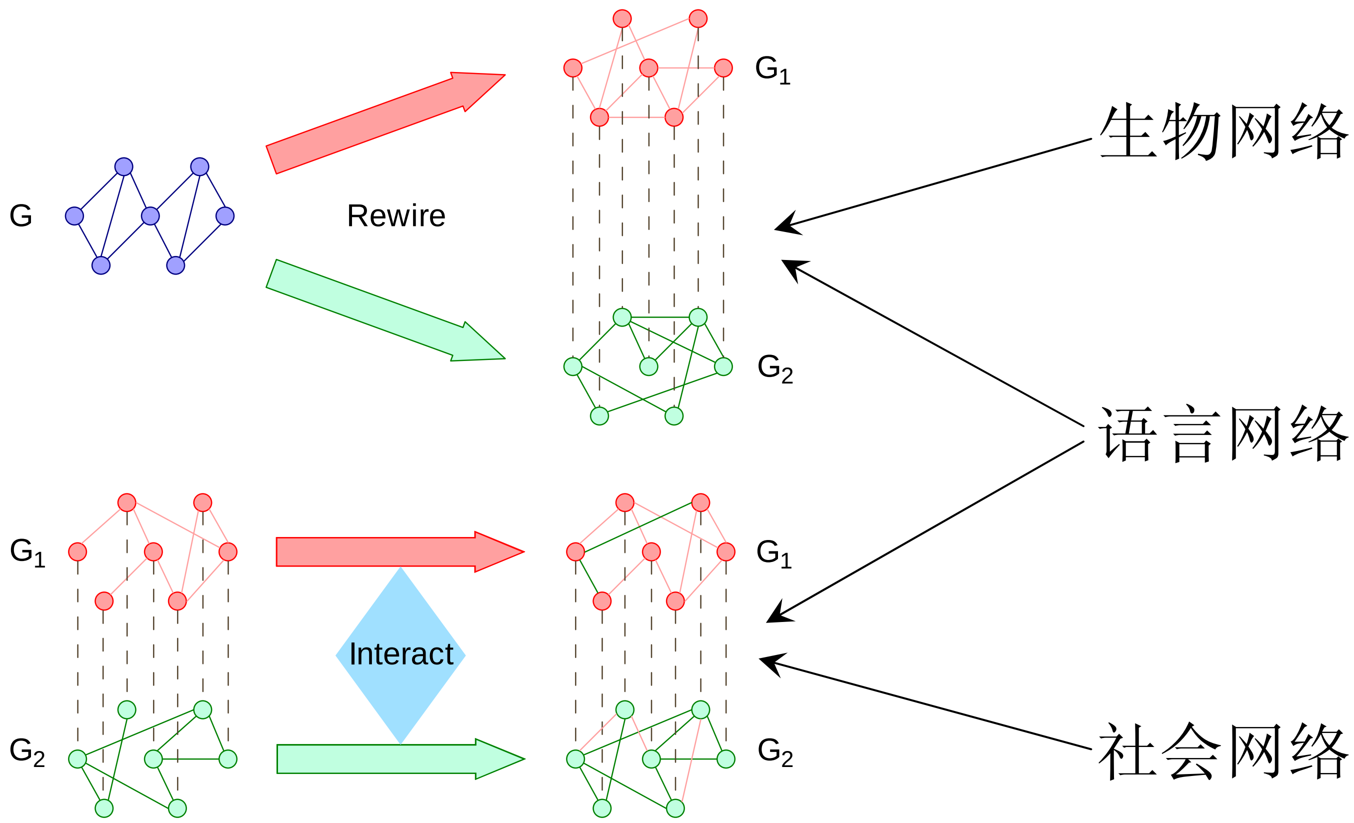


FIG. 2. (Color online) The numbers of the total QQ friends as well as the numbers of the QQ friends with telephone numbers for the 24 investigated QQ users. It is revealed that the ratio of the number of the QQ friends with telephone numbers to the number of the total QQ friends for all the investigated 24 QQ users is equal to 1574/2702=0.58.

就总体样本数据而言，两个个体如果在QQ网络中是好友，那么他们在电话网络中是好友的可能性为58%，参考文献一，反之尚未进行进一步验证。

复杂网络的协同演化的两种模式



结构信息的局限性

- **局限一：**在单个网络中，通常存在大量的节点具有完全相同的局部结构特征，比如相同的度值，相同的聚类系数等等。
- **局限二：**匹配网络并非完全等价，即个体在不同网络中只是一种概率相似，通常可能具有不同的局部特征。
- 这样的局限性决定了仅对节点按某些局部特征进行排序，并以此来确定不同网络之间的节点匹配注定不会取得理想的匹配效果。
- **比较科学的做法：**通过其他方法获知某些匹配节点对，结合这些已匹配节点对以及网络结构提供的信息设计节点匹配算法。

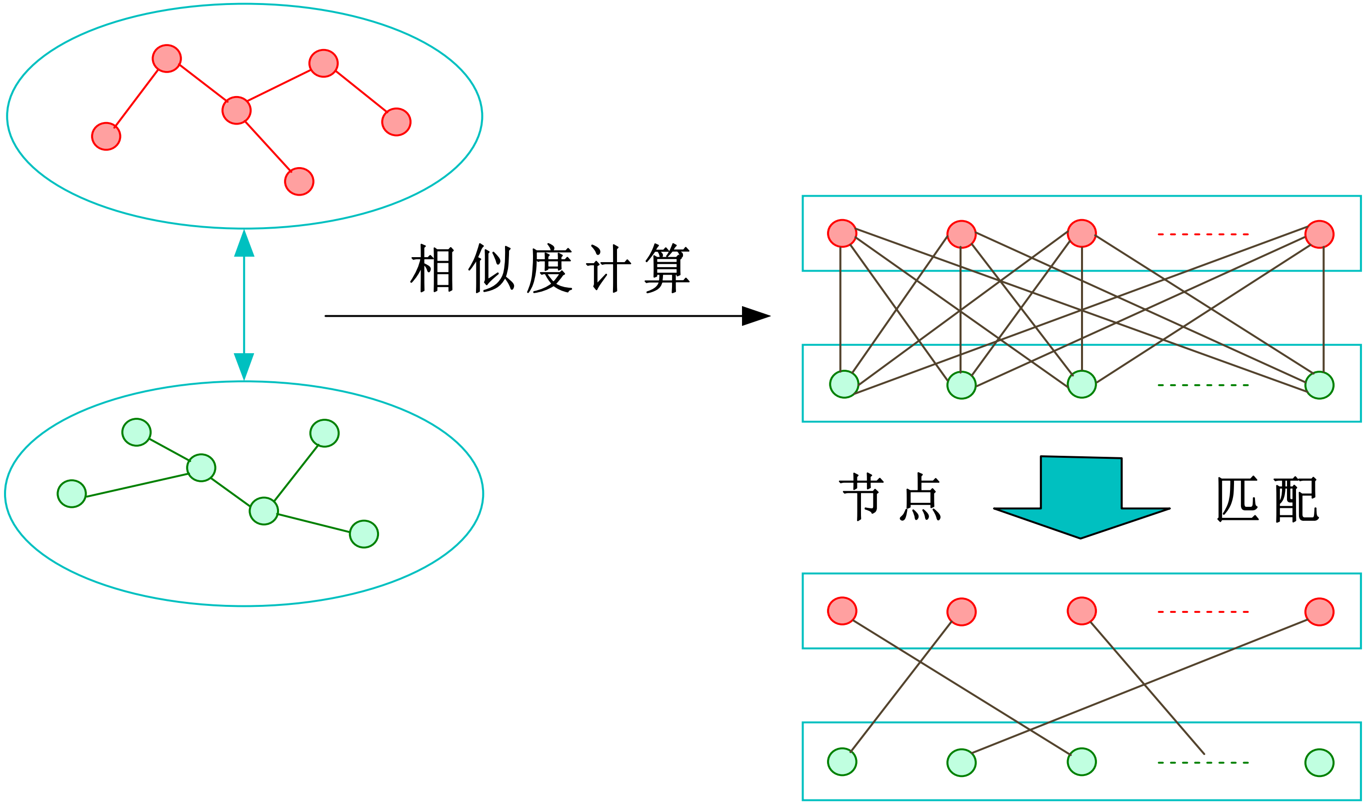
已匹配节点对的获取（社会网络为例）

- **途径一：**在因特网上注册账号时，有一部分用户可能会留下自己的电邮地址或电话号码。

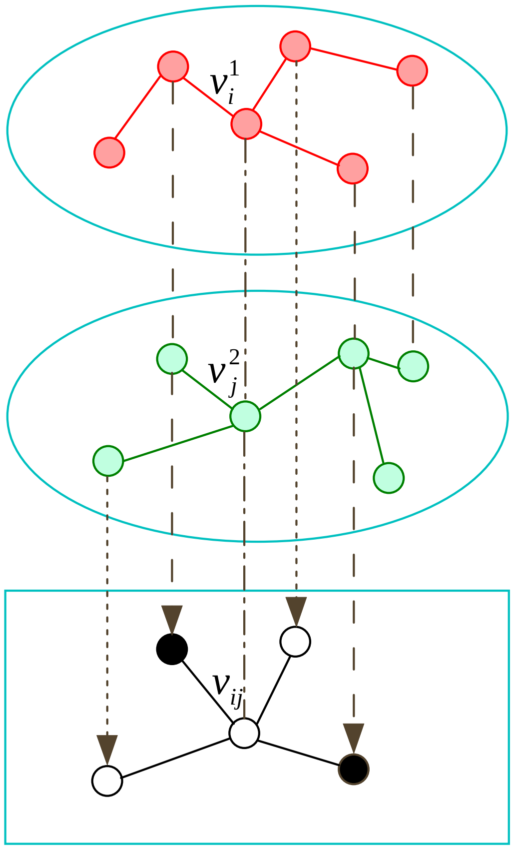
- **途径二：**某些知名网络社区提供的附件服务，比如QQ网络已推出自己的专用电子邮箱（每个QQ号自动分配一个邮箱），从而形成QQ网络和Email网络之间的“已匹配节点对”。

- **途径三：**账号绑定功能的推行以及网络实名制的倡导。

初步匹配算法框架



相似度计算



$$S(v_i^1, v_i^2) = \frac{n_M(v_i^1, v_i^2)}{n_L(v_i^1) + n_L(v_i^2) - n_M(v_i^1, v_i^2)}$$

分子：两节点共同连接的已匹配节点对数。

分母：两节点共同的邻居数。

《图论》中加权二分图最大匹配算法

- Bellman-Ford algorithm—算法速度为O(N^4)
- Dijkstra algorithm—仅针对非负权重，算法速度为O(N^3)
- Kuhn-Munkres algorithm（Hungarian algorithm）—组合算法，算法速度为O(N^3)

PHYSICAL REVIEW E **80**, 026103 (2009)

Node matching between complex networks

Qi Xuan* and Tie-Jun Wu

Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

(Received 10 November 2008; revised manuscript received 26 March 2009; published 4 August 2009)

Revealing corresponding identities of a dedicated individual in several different complex systems is a common task in many areas, and this task is transferred to a node matching problem among complex networks in this paper. A feasible node matching algorithm based on network structure is proposed. Through solving node matching problems on different types of networks by our algorithm, it is revealed that the structure of the networks under study may significantly influence the final matching results. For example, it is found that higher matching precision can be obtained on random networks with moderate density of links, and the results on small-world networks are always better than those on random or regular networks. Moreover, in scale-free networks, it seems that hub nodes play dominant roles, i.e., better matching results can be expected by selecting nodes with larger degrees as the revealed matched nodes. These findings will help us design more efficient node matching algorithm in the future.

DOI: 10.1103/PhysRevE.80.026103 PACS number(s): 89.75.Hc, 89.65.Ef, 89.70.Eg, 89.75.Kd

随机网络匹配结果

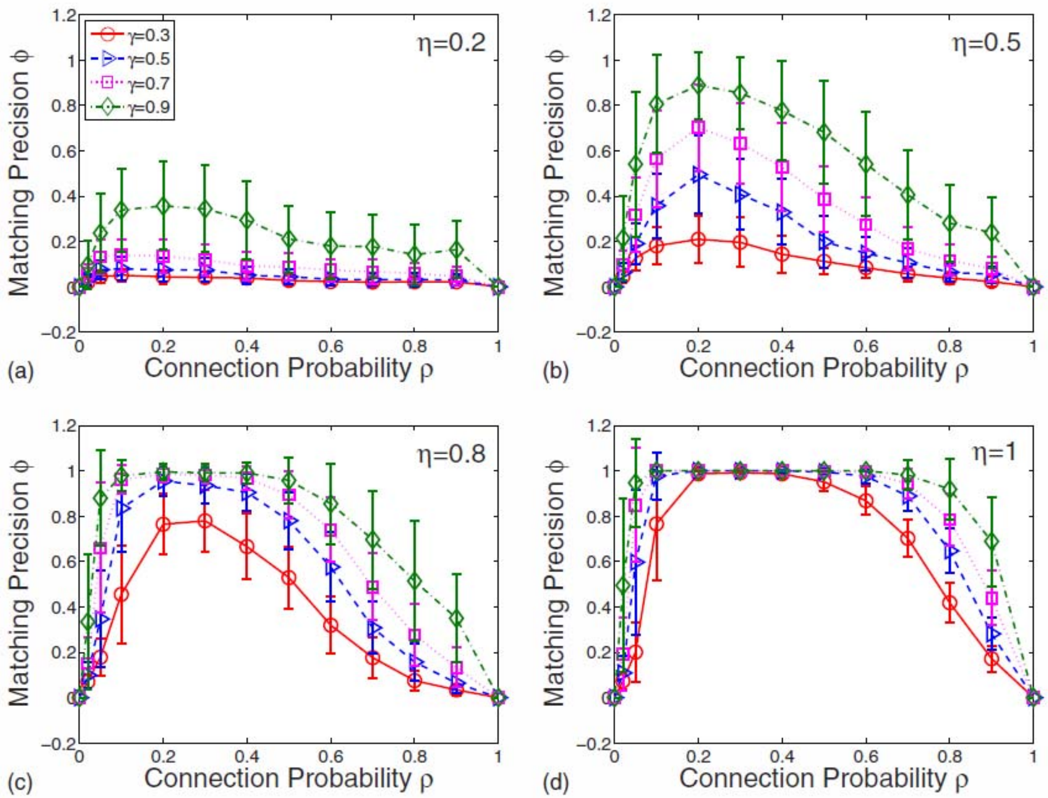


FIG. 4. (Color online) The relationships between the matching precision ϕ and the connection probability ρ for various sample ratios γ=0.3,0.5,0.7,0.9 and different interactional degrees η=0.2,0.5,0.8,1. For each parameter set $\{\rho, \eta, \gamma\}$, the experiment is implemented on 100 different pairs of random networks, each of which has N=100 nodes.

小世界网络匹配结果

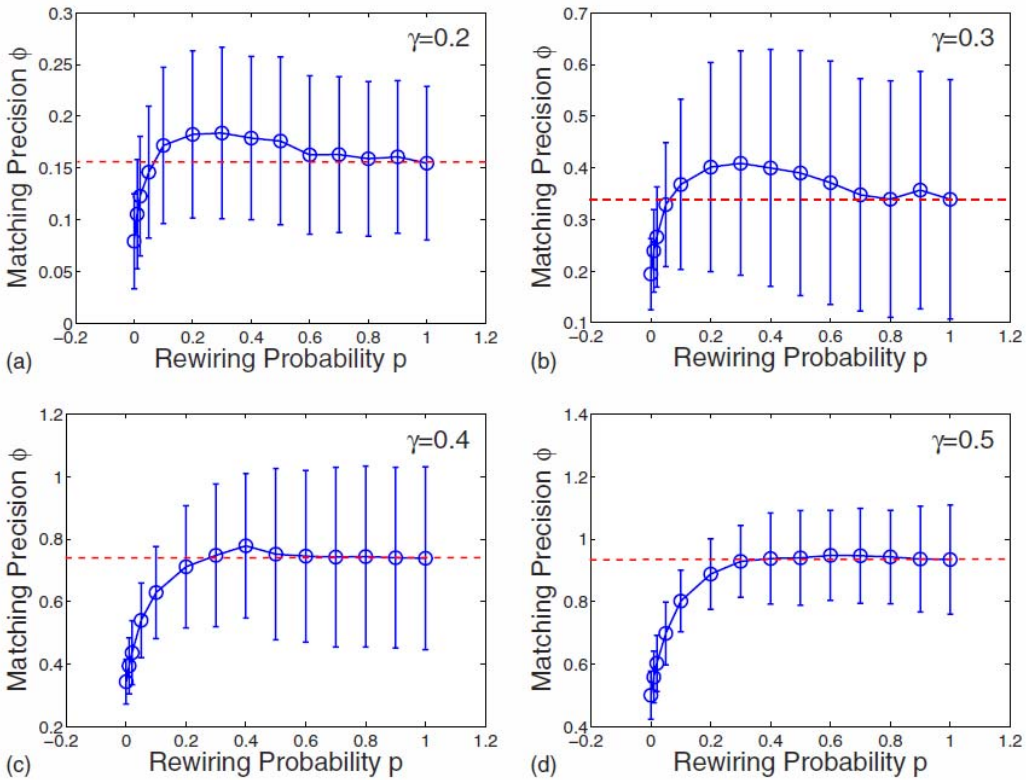


FIG. 7. (Color online) The relationships between the matching precision ϕ and the rewiring probability p for small-world networks under various sample ratios γ=0.2,0.3,0.4,0.5. For each parameter set $\{p,\gamma\}$, the experiment is implemented on 1000 different pairs of small-world networks, each of which has N=100 nodes and an average degree $\langle k \rangle$=10. The red (dark) dashed lines denote the matching precision of the algorithm on the random networks when p=1.

无标度网络匹配结果

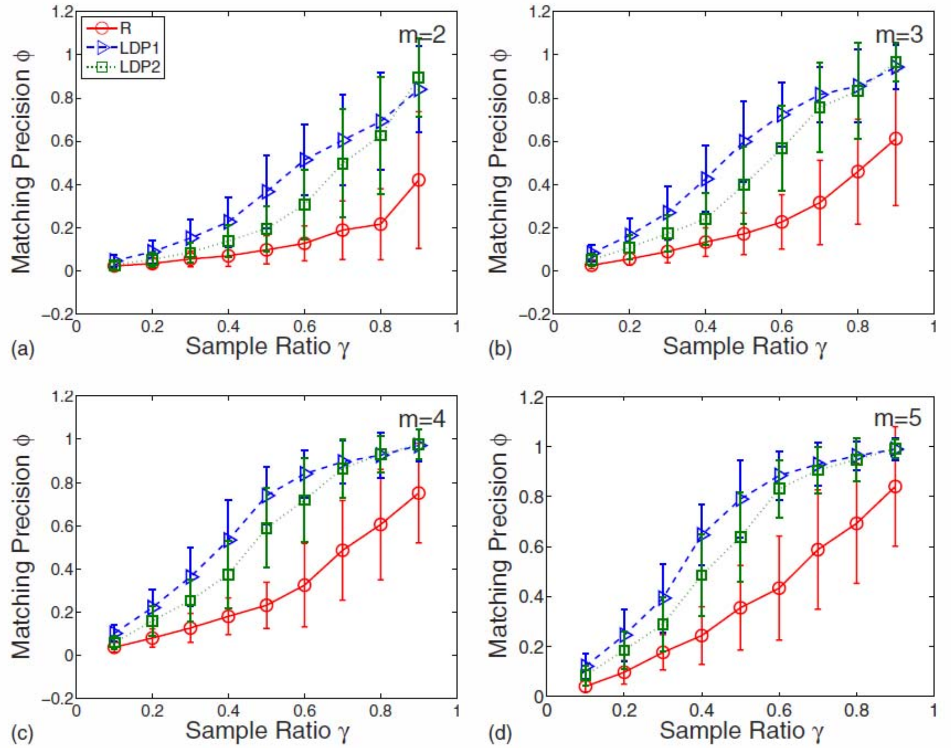
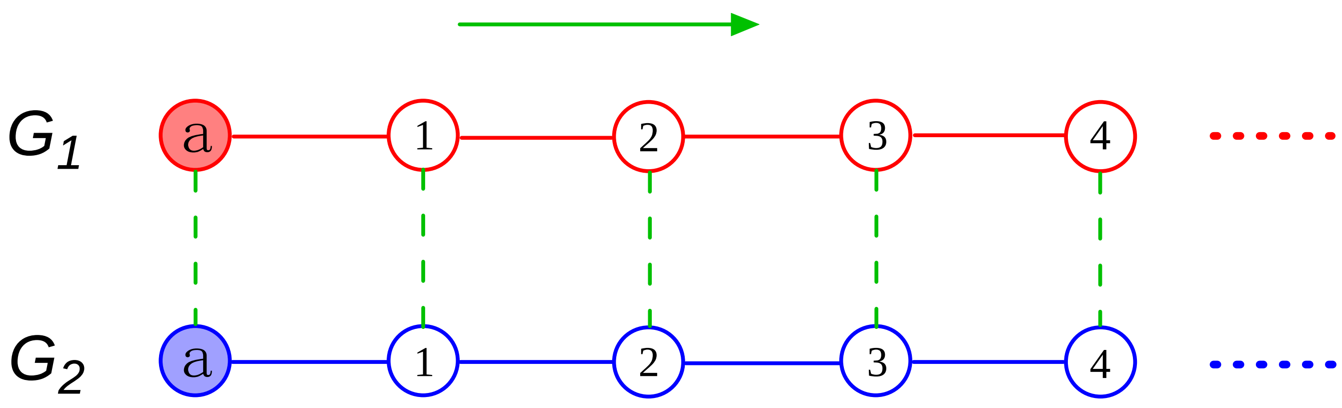
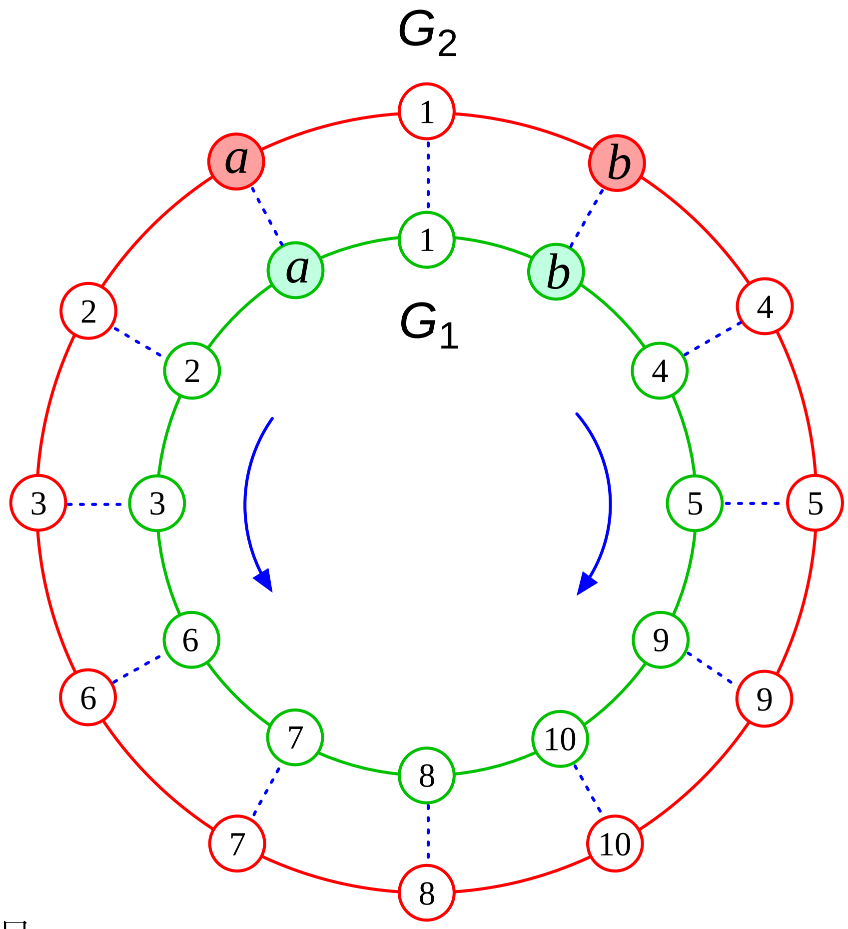


FIG. 8. (Color online) The relationships between the matching precision ϕ and the sample ratio γ by adopting the three different revealed matched nodes selection rules, i.e., R, LDP1, and LDP2, for scale-free networks with different parameters. (a) With parameters $N=100$, $m=m_0=2$. (b) With parameters $N=100$, $m=m_0=3$. (c) With parameters $N=100$, $m=m_0=4$. (d) With parameters $N=100$, $m=m_0=5$. For each parameter set $\{\gamma, m, m_0\}$ and each selection rule, the experiment is implemented on 100 different pairs of scale-free networks. It is shown that, generally, better matching results can be derived by selecting revealed matched nodes following the degree based rules LDP1 and LDP2 than following the random selection rule R. Moreover, in most cases, LDP1 is prior to LDP2 when $\eta_1 > \eta_2$ in this experiment.

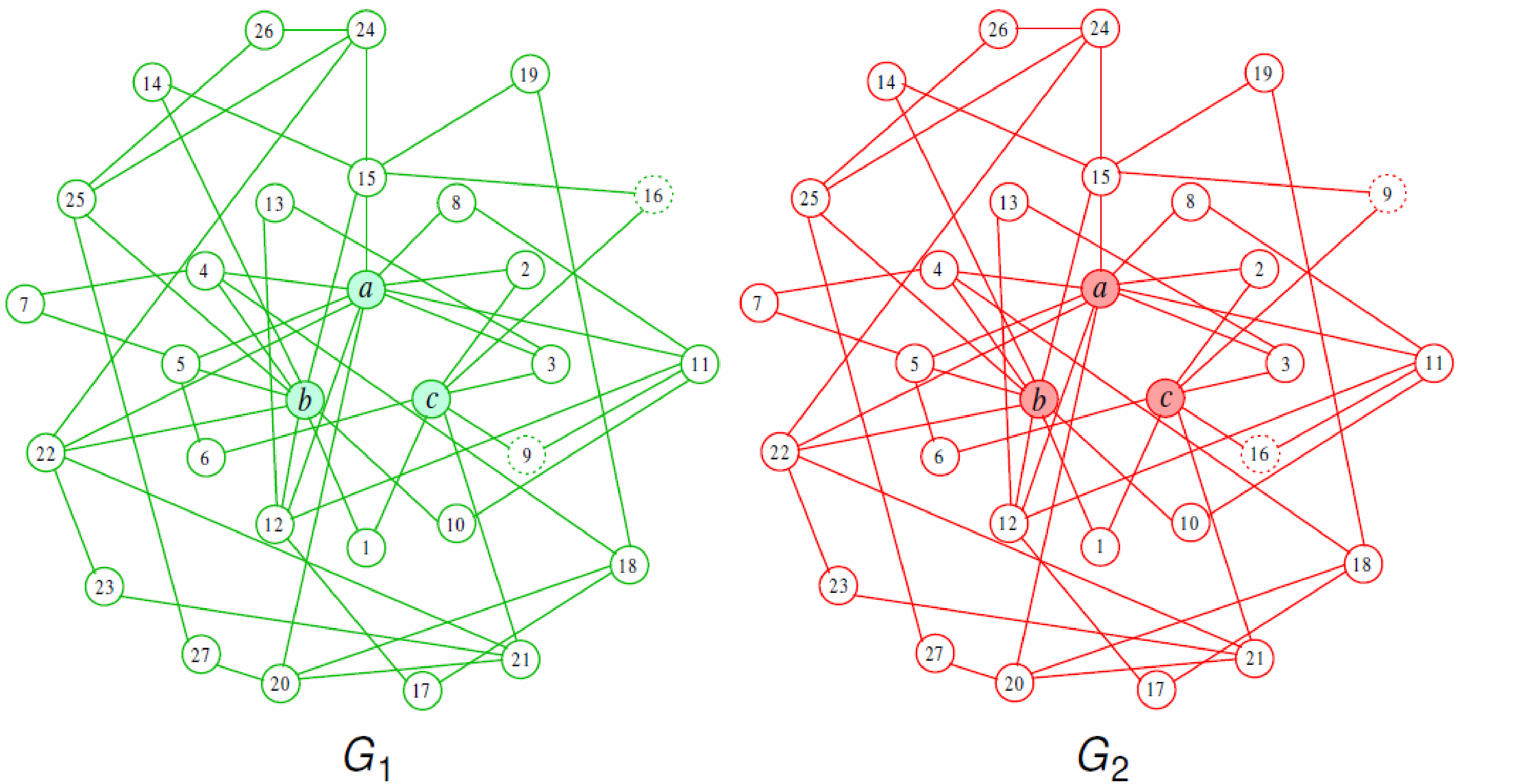
迭代匹配算法（实例一）



迭代匹配算法（实例二）

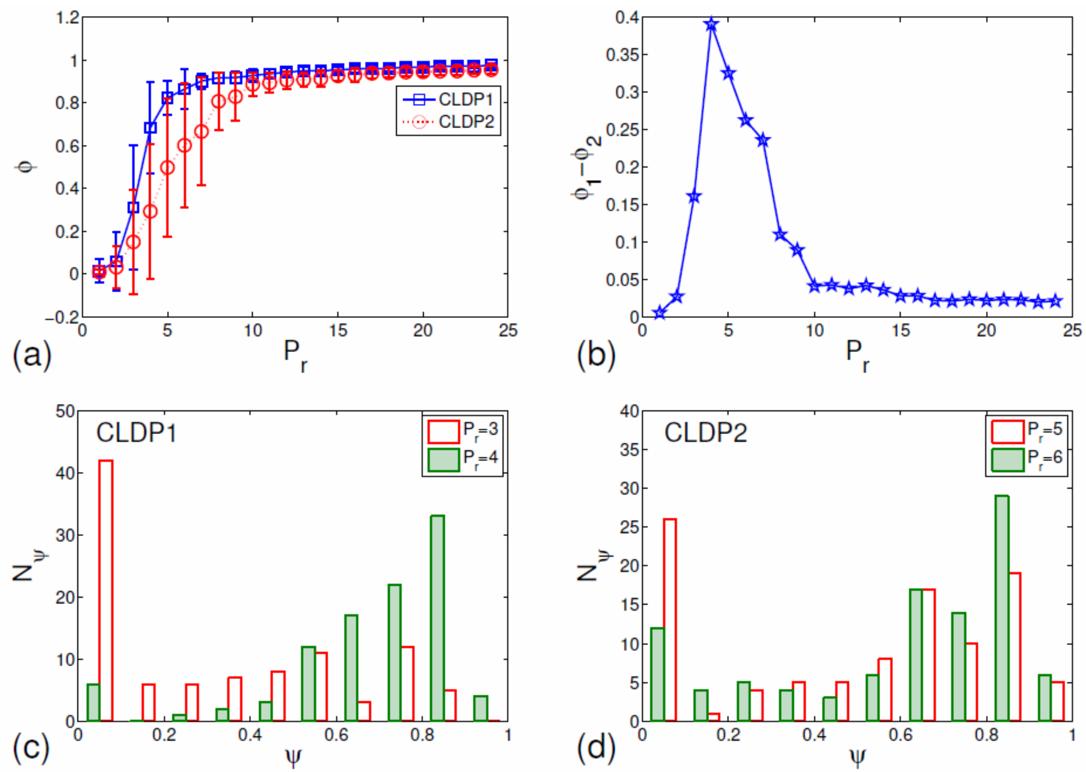


迭代匹配算法（实例三）



迭代匹配算法相关文章

Iterative node matching between complex networks. Submitted.

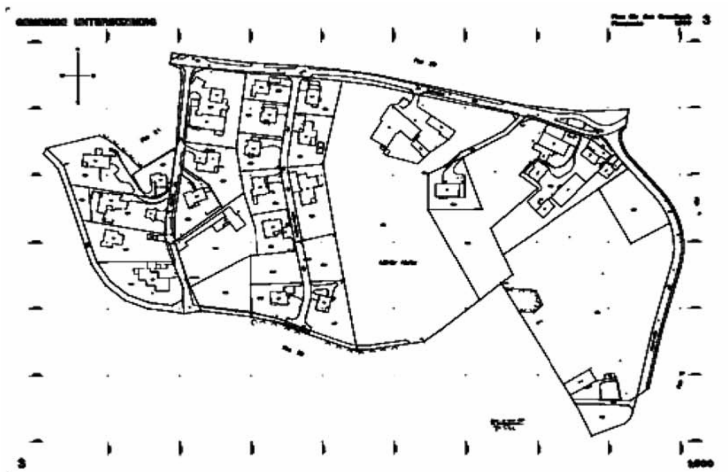


算法速度

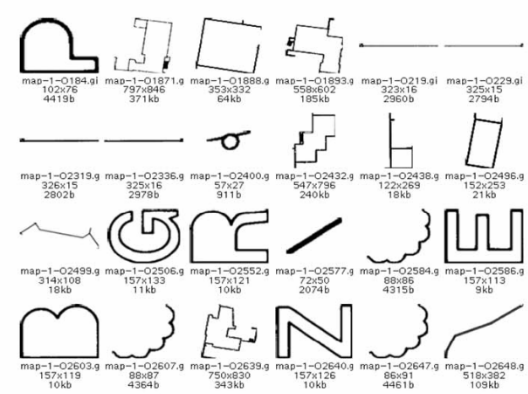
最坏情况：O(N*ln(N))

最好情况：O(N)

相关话题：人工智能中的子图匹配



Finding



不同之处：

节点匹配：
追求点对点匹配
提供已匹配节点对

子图匹配：
只寻求与目标结构相同的子图
不需已匹配节点对


谢谢！