社会计算

第二节

网络链接与符号性

徐童　　2025.3.2

课前提醒

- **友情提示：可以开始组队了**

- 本课程的任务包括两次实验与一份课程报告
- 所有任务均由不多于 8 人一组共同完成
- 所有小组成员得分相同，不再区分

- 如果还未找到队友，可以利用课程群
  - QQ群：1035623172，欢迎各位积极发言
  - 注意：每次任务可以重新分组！

课前提醒

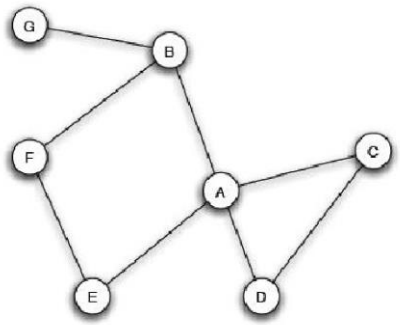
- **第二个实验内容：社交游戏设计**

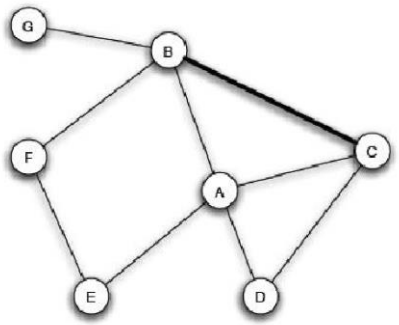
- 核心任务：设计一款社交游戏并进行三轮实验
- 要求：
  - 1）玩家应扮演不同身份并具有相应分工，在游戏过程中可根据需要选择是否暴露身份
  - 2）玩家之间应有社交互动与合作、竞争行为，应有明确收益目标及利益分配过程
  - 3）玩家之间仅通过文字进行交流（排除场外因素干扰），允许在交流中说谎
  - 特殊要求：大模型的参与
    - 游戏共进行三轮，第一轮全人类参加，第二轮GM挑选部分玩家替换为大模型，第三轮为纯大模型进行。大模型参与游戏的形式自定，不做统一要求。

上节回顾

- **三元闭包与强/弱关系**

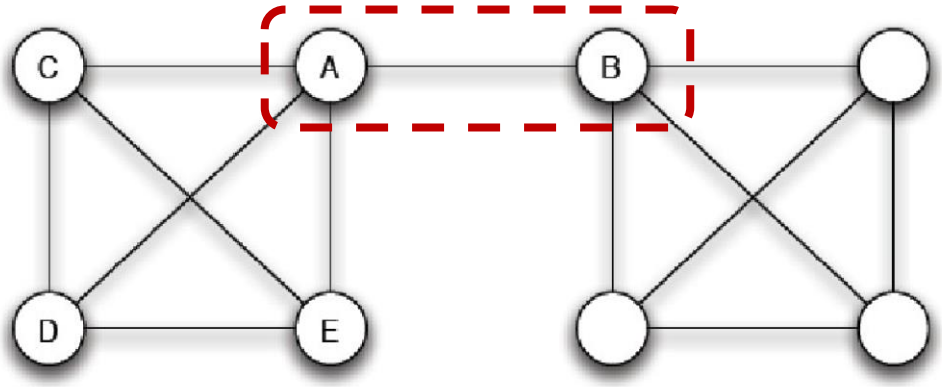
- 强关系 V.S. 弱关系
- “三元闭包”的形成与强三元闭包原理
  - 新生成的边可能与已有的边是同质的
  - 但存在另一种可能，新生成的边是在已有的边影响下“被动”形成的
    - 已有边/共同好友的“中介”作用
  - 强三元闭包原理
    - 中介的两个强关系好友间必有关系（无论强弱）



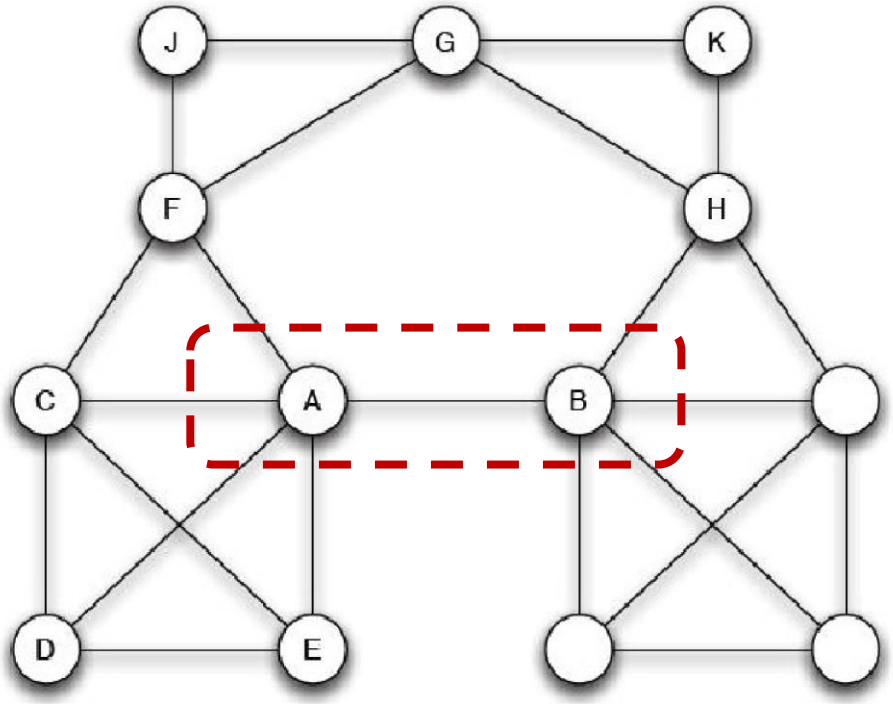


- **两类弱关系：桥与捷径**

移除了“桥”，则原图将被拆为若干连通子图



移除“捷径”并不会破坏连通性，
但节点间的通讯代价将上升



上节回顾

- **小世界现象与W-S模型**

- 六度空间理论与相关实验
  - 在Milgram实验之前，人们意识到世界很小，但没有切实的证据
  - Milgran实验至少证明了“小世界现象”对应的两个惊奇
    - ✓ 世界是小的：社会网络中包含着丰富的短路径
    - ✓ “有意识的转发”能够“自动”找到这些短路径

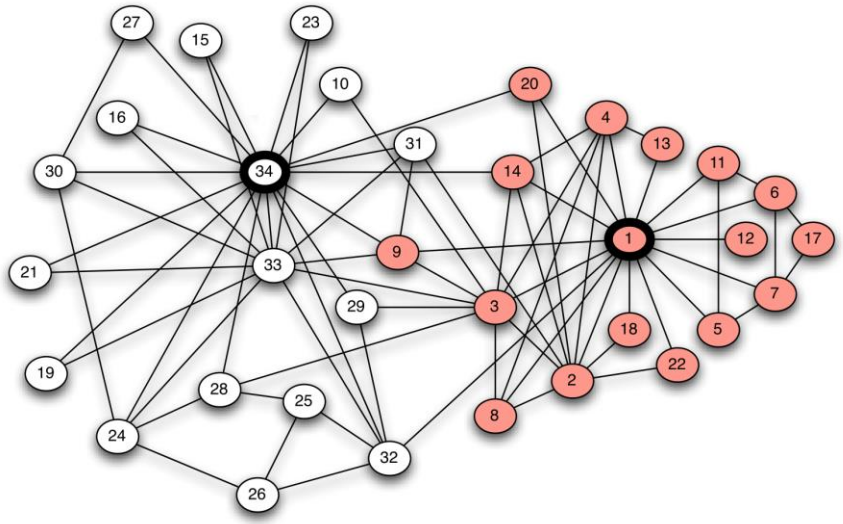
➢ W-S / W-S-K模型，为模拟小世界现象提供了量化模型

- 注意其中有关远程路径的设计与选择问题

本节主旨

- **网络中的“边”从何而来?**

- 我们所见的社会网络，是人与人之间复杂关系投影于二维空间的写照
  - 除了先天存在的关系，其他的关系从何而来？又蕴含着何种规律?
  - 如何衡量友好 – 敌对的关系符号性？这种符号性又是如何维持平衡的?

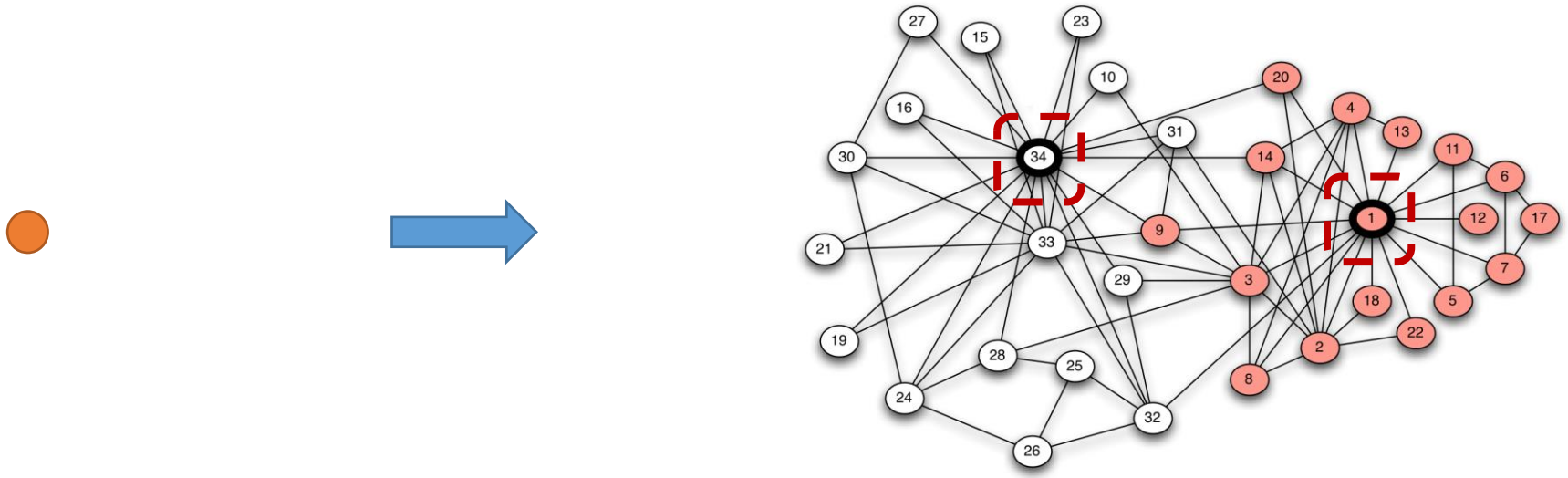


本节目录

- **网络生成模型**
  - 随机网络
  - 幂律特性与B-A模型
- 链接预测模型
  - 基于邻居关系的模型
  - 同质性与链接选择
- 网络符号性

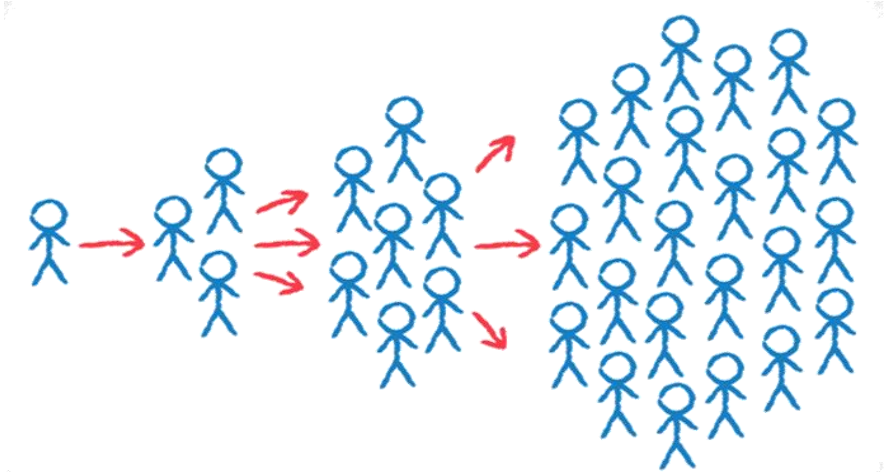
• 从无到有的网络生成

- 我们先考虑一类特殊的网络生成问题：从无到有的网络
  - 如何从初始的单一节点，扩散为一个颇具规模的社会网络?
  - 仔细观察网络中的“中心节点”



从无到有的网络生成

- 一些现象或许能够帮助我们理解这一网络生成的过程
  - （1） Alex Pentland教授的红气球“传销”
    - “据称，Pentland在短短数小时内便动态组建了一支成员多达5000人的团队，这5000名队员中的每个人又平均通知了400名**朋友**”
    - “不仅奖励正确告知气球地点的人，还奖励那些把找到气球的人成功介绍给团队的人”



从无到有的网络生成

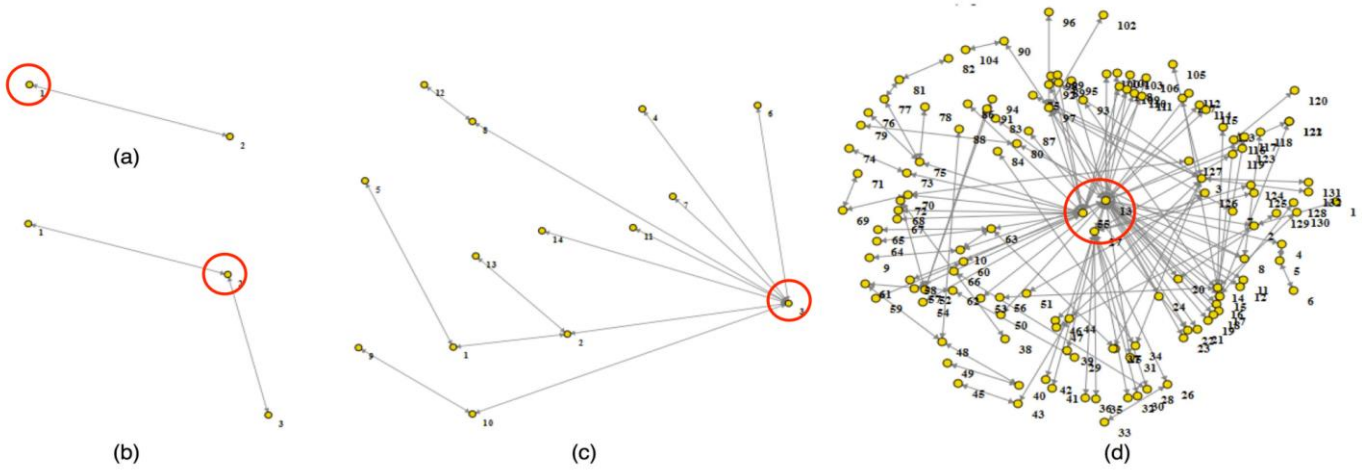
- 一些现象或许能够帮助我们理解这一网络生成的过程
  - （2） Meetup 中用户参与活动的动机分析
    - “虽然相比于拒绝邀请的用户，那些活动的参与者体现出更强的兴趣，然而有趣的是，长期参加活动的核心用户并没有比那些偶尔参加活动的人体现出更为浓厚的兴趣”
    - “一种合理的猜测是，这些人并不是因为兴趣而参加活动，更可能是因为其他朋友的邀请而出席”

Table 2: Comparison for user preference to events.

	Single Event		All Events
Attendant	0.108	Active Users	0.106
Absentee	0.094	Overall Users	0.105
P-Value	0.016	P-Value	0.334

• 从无到有的网络生成

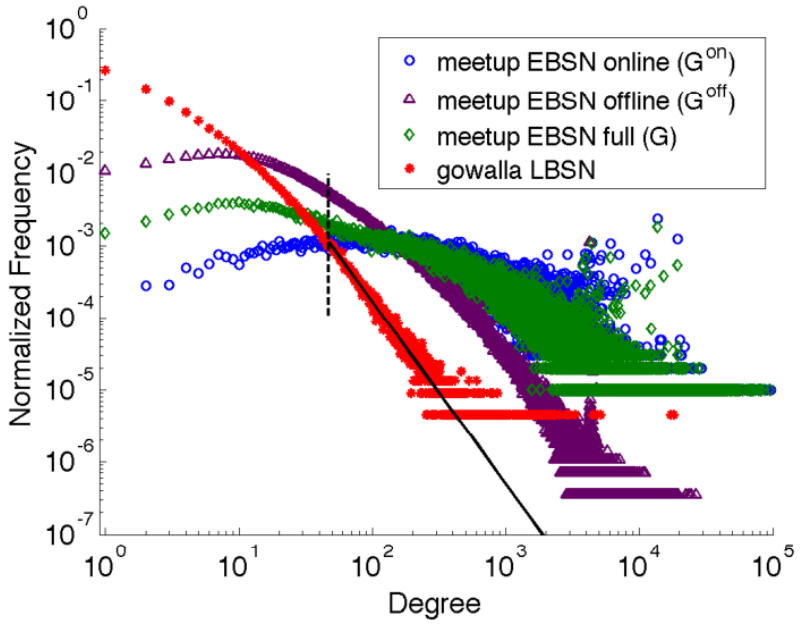
- 由此，我们或许可以总结出“从无到有”的网络生成过程
  - 核心节点发起，先吸引熟悉的人加入网络，这些人再吸引更多的人，逐渐“滚雪球”
  - 此类网络生成的过程，往往具有明显的“星形网络”特性



Chunyan Wang, et al., From Face-to-Face Gathering To Social Structure, CIKM-2012

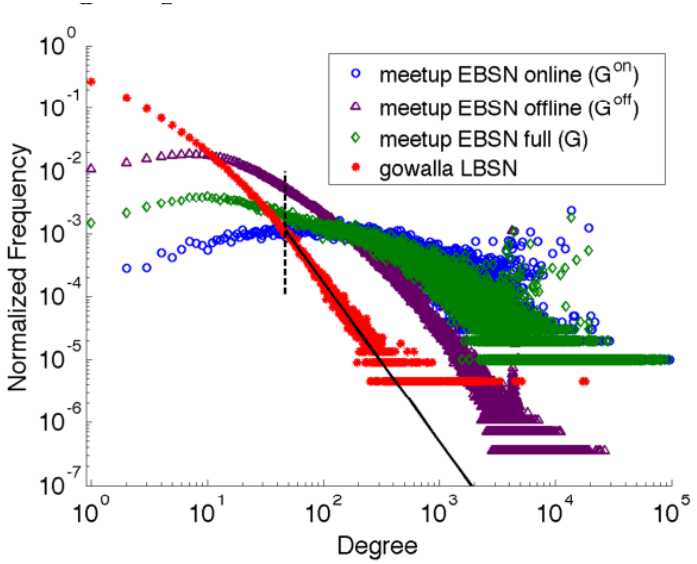
- **从无到有的网络生成**

- 拓展小知识：从无到有的星形网络中，幂律现象会更为明显
  - 相比于一般网络近似为幂律分布，从无到有的EBSN（事件驱动的社会网络）往往呈现更强的重尾分布现象
  - 一种合理的解释是：这一类网络衍生于熟人“传销”，位于核心地位的节点引入了大量链接，而被引入的节点之间则很少存在关联
  - 事实上，人们也同时观测到，随着这一类网络规模的不断增长，网络中的边反而越来越稀疏



• 从无到有的网络生成

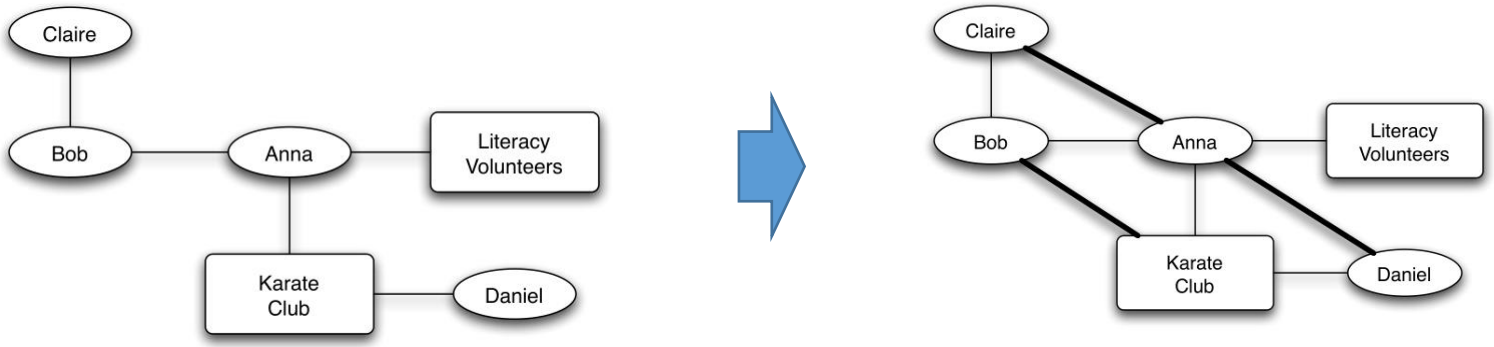
- 一些题外话：在现实的社会网络中，“星形网络”特性似乎并不明显？
  - 事件驱动的社会网络，由于用户对于新事件/环境的陌生，而更依赖已有社交关系
  - 一般社会网络中，用户“先天地”已经具备了较为稠密的社交关系，尤其是弱连接



Chunyan Wang, et al., From Face-to-Face Gathering To Social Structure, CIKM-2012

更为一般化的网络生成

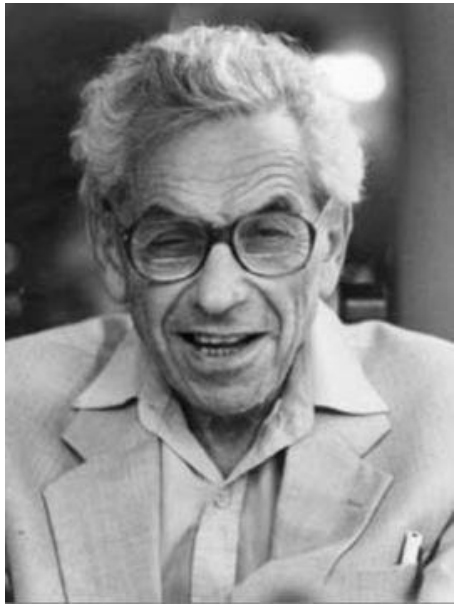
- 从无到有的网络并不具有一般性
  - 从无到有的网络生成仅能描绘一些小范围、新生成网络的增长过程
  - 通常情况下，网络中的绝大部分节点是已经存在的，需要增长的只是他们之间的关系
  - 如何描述这种链接的生成过程?



更为一般化的网络生成

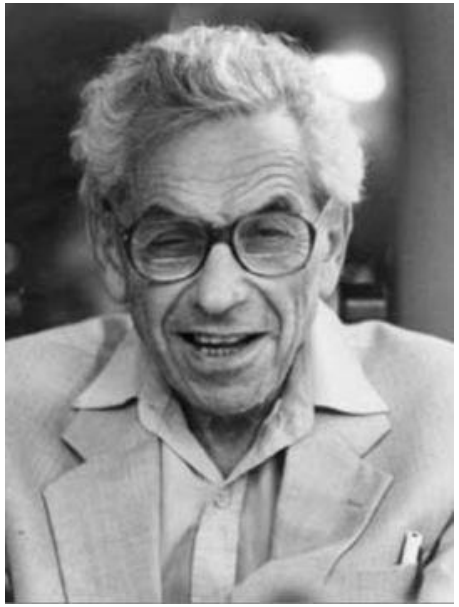
- Erdos-Renyi的随机图理论
  - 假设边的生成过程相互独立
  - 边的生成过程按照以下形式描述
    - 网络中一共有 N 个节点，随机从中生成若干条边
    - 每条边生成的概率为 p
    - 则生成一张边数为 n 的图的概率为

$$P(G) = p^n \cdot (1-p)^{\frac{N(N-1)}{2}-n}$$



• 更为一般化的网络生成

- Erdos-Renyi的随机图理论
  - 随机图理论简单易行，也能够在一定程度上反应人们的认知
    - ➢ 人与人的相识起源于奇妙的邂逅，而这种邂逅包含偶然（🐱）
  - 然而，该模型由于过于简化，以至于无法描述真实情况
    - 首先，该模型无法描述小世界网络
    - 其次，该模型很难有自发形成的群组
    - 更为重要的是，该模型不符合幂律分布

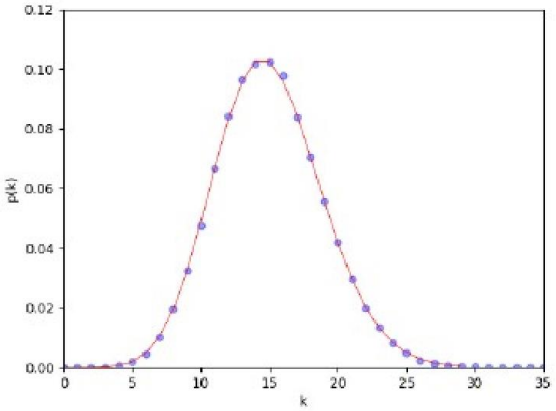


更为一般化的网络生成

- Erdos-Renyi的随机图理论
  - 在随机图理论中，节点的度可以采用以下公式进行估计
    - 节点的度为 k 的概率为

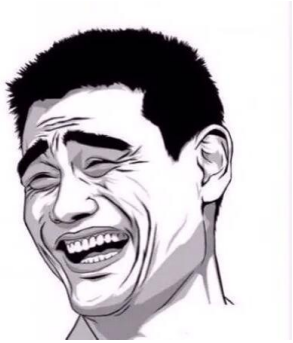
$$p(k) = \binom{N-1}{K} P^k (1-P)^{N-1-k}$$

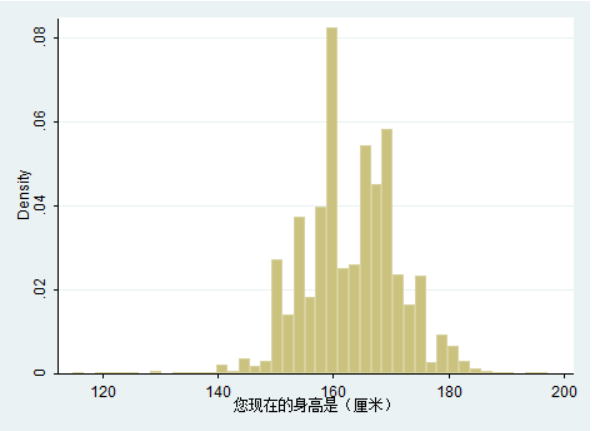
  - 由此可知，随机图理论生成的网络，其度的分布大致遵循泊松分布
    - 其分布图如右图所示 →



- **更真实的幂律网络生成**

- 为什么泊松分布不合理?
  - 泊松分布适合于描述单位时间（或空间）内随机事件发生的次数
  - 现实生活中，确实有很多现象可以用泊松 / 正态分布近似描述
    - 例如，很难见到特别高和特别矮的人
    - 在这一类分布中，平均值往往表示“典型行为”





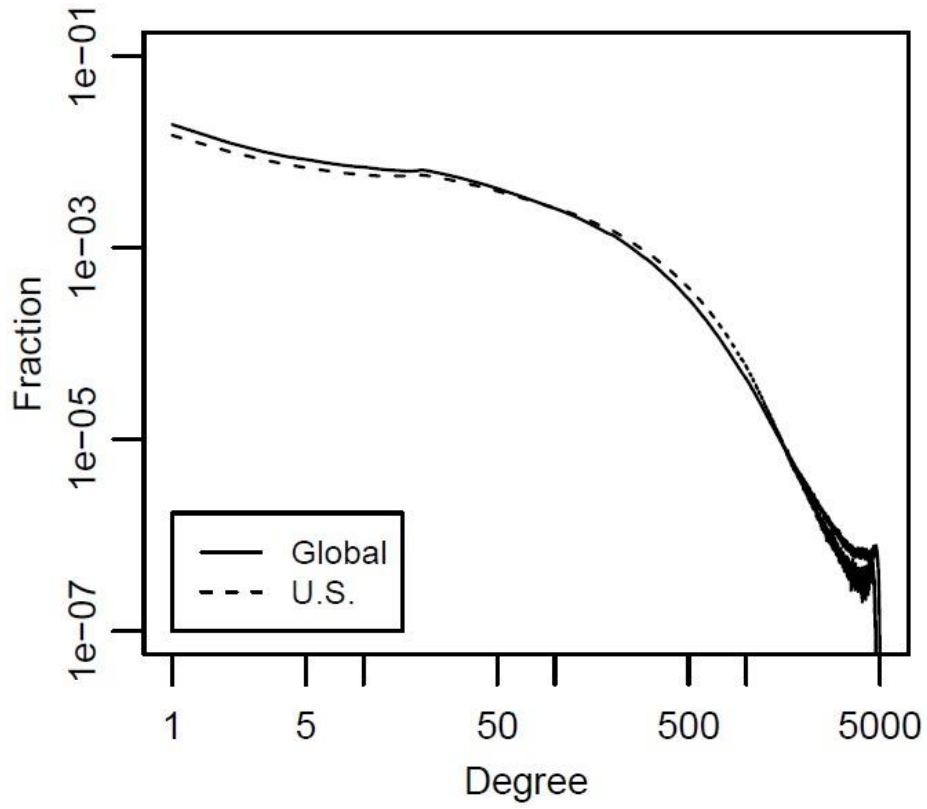
- **更真实的幂律网络生成**

- 为什么泊松分布不合理?
  - 但在另一些场景中，平均数并不是一个好指标
    - 一个典型的例子：2020年全国居民人均可支配收入32189元

先前回顾

- **更真实的幂律网络生成**

- 我们知道，真实网络中的节点度数往往符合**幂律**（Power-law）分布
  - 总结：少数节点拥有大多数的边
  - 这些少数节点即形成了所谓“影响力节点”，也就是俗称的“大V”
  - 基于真实的网页链入数据显示
    - 拥有 k 个入边的网页所占比重，近似与 $1/k^2$ 成正比



- **更真实的幂律网络生成**

- 幂律的威力，并不限于“被平均”

凡有的，还要加倍给他叫他多余；没有的，连他所有的也要夺过来

——《新约·马太福音》

- 富者愈富现象：领先者基于其累积优势而更快、更容易地获得成功
  - 如，微博某女王 23 个月粉丝破千万
    - 300万粉丝增长至500万粉丝，历时4个月；
    - 500万粉丝增长至700万粉丝，历时3个月；
    - 700万粉丝增长至900万粉丝，仅历时2个月……

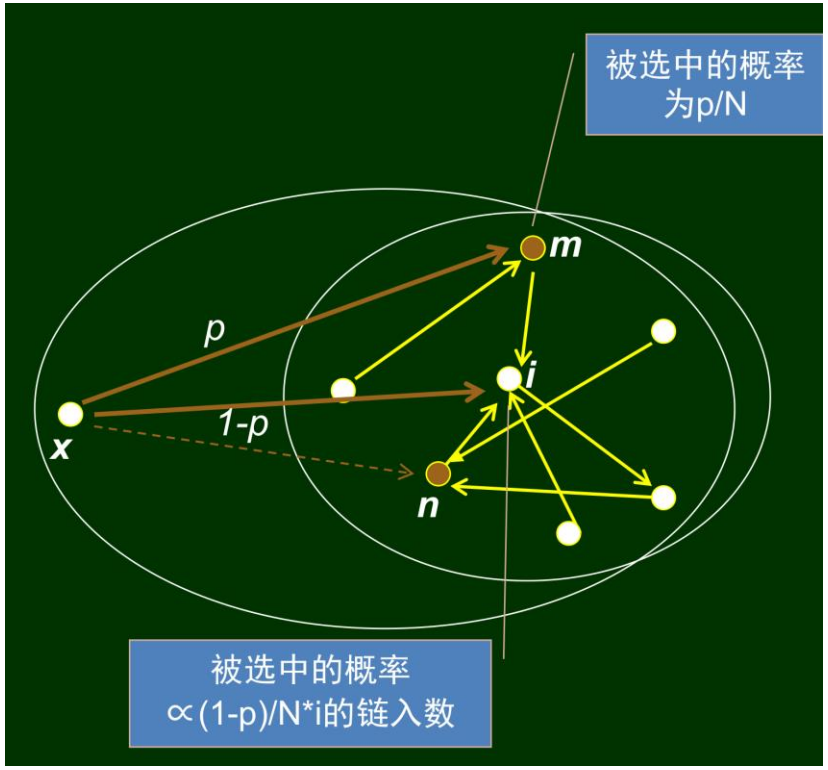
- **更真实的幂律网络生成**

- 为什么网络中呈现幂律分布？一种启发式的网络生成方式如下：

1. 网页按顺序创建，编号为1, 2, 3, …, N
2. 当网页 j 被创建时，按照概率 p 选择以下（a）或者（b）的操作方式产生链接
   a) 以概率 p，均匀随机地选择一个早先创建的网页 i，建立一条到 i 的链接
   b) 以概率 $1-p$，均匀随机地选择一个早先创建的网页 i，建立一条到 i 指向的网页的链接

- **更真实的幂律网络生成**

- 在前述启发式网络生成流程下，节点链接数增长率与其当前链接数成正比：



- **更真实的幂律网络生成**

- 上文中的(b)操作，实际上是人们从众的一种体现，体现了“择优连接”的思想
  - 大家的选择（People’s Choice）总是更可靠的
  - 例如，当你在淘宝上购物的时候，你是否会参考商品的已成交量 / 评论数?

更真实的幂律网络生成

- 基于幂律分布启发的Baraba'si-Albert模型
  - 假设已经存在了一个一定规模的网络
  - 每次引入一个新节点，并引入 m 条边（$m \le N$，N为当前已有节点数）
  - 则网络中某个节点 i 吸引一条新边的概率与其度数 k_i 相关，公式如下：

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}.$$

    - 显然，每个节点吸引新边的能力是不同的，度数越高，吸引新边的能力越强
    - 因此，这一公式也被视作体现了“择优连接”（Preferential Attachment）的思想

- **更真实的幂律网络生成**

- 基于幂律分布启发的Baraba'si-Albert模型
  - Baraba‘si-Albert模型中，单个节点的度数演化规律：

$$\sum_j k_j = 2mt$$

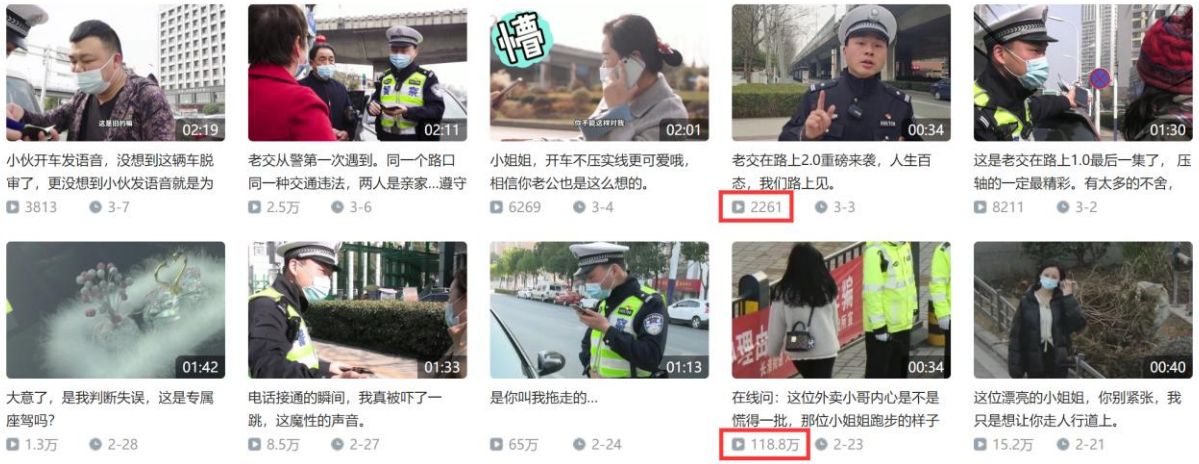
$$\begin{aligned}\frac{\partial k_i}{\partial t} &= m\Pi(k_i) \\ &= m\frac{k_i}{\sum_j k_j} \\ &= m\frac{k_i}{2mt} \\ &= \frac{k_i}{2t}\end{aligned}$$

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{0.5}$$

其中，t 为当前时刻，t_i 为节点 i 加入网络的时刻

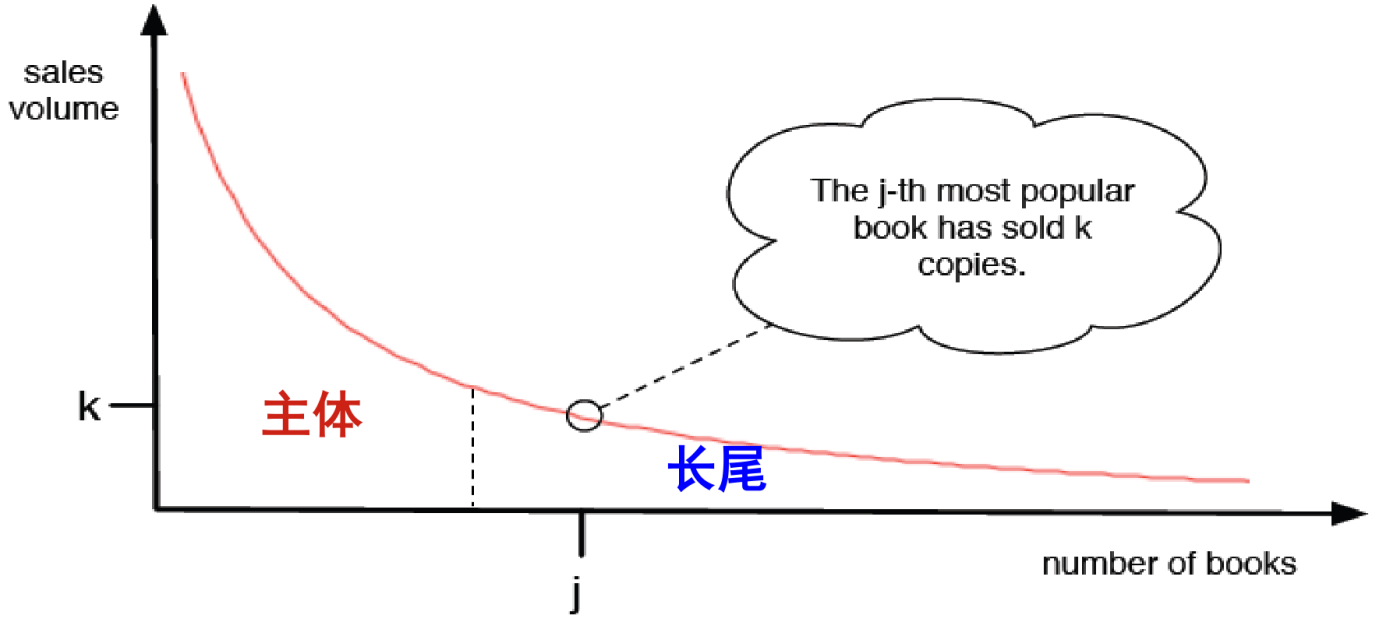
• 更真实的幂律网络生成

- 一个有趣的现象：富者更富现象的“不确定性”
  - 原始积累期存在着很强的不确定性和初始波动性，取决于小部分人的选择决定
  - 一但“富”到了一定的程度就开始“起飞”，此时趋势就难以逆转了
    - 质量较差的内容可能反而能够胜出，仅仅因为它比竞争对手更快达到临界值



- **更真实的幂律网络生成**

- 最后想说的话：与大V相对应的，是“被平均”的大多数
  - 长尾效应：少数头部垄断了大多数的信息/关注，而零散的、差异化的、小众的信息/关注会形成长长的“尾巴”



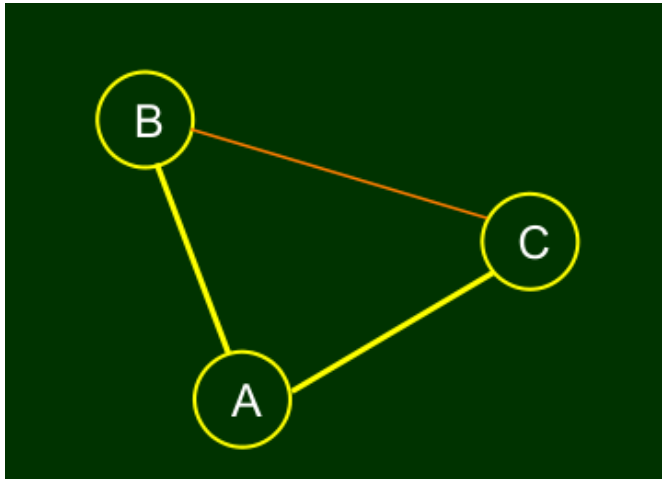
本节目录

- 网络生成模型
  - 随机网络
  - 幂律特性与B-A模型
- **链接预测模型**
  - 基于邻居关系的模型
  - 同质性与链接选择
- 网络符号性

先前回顾

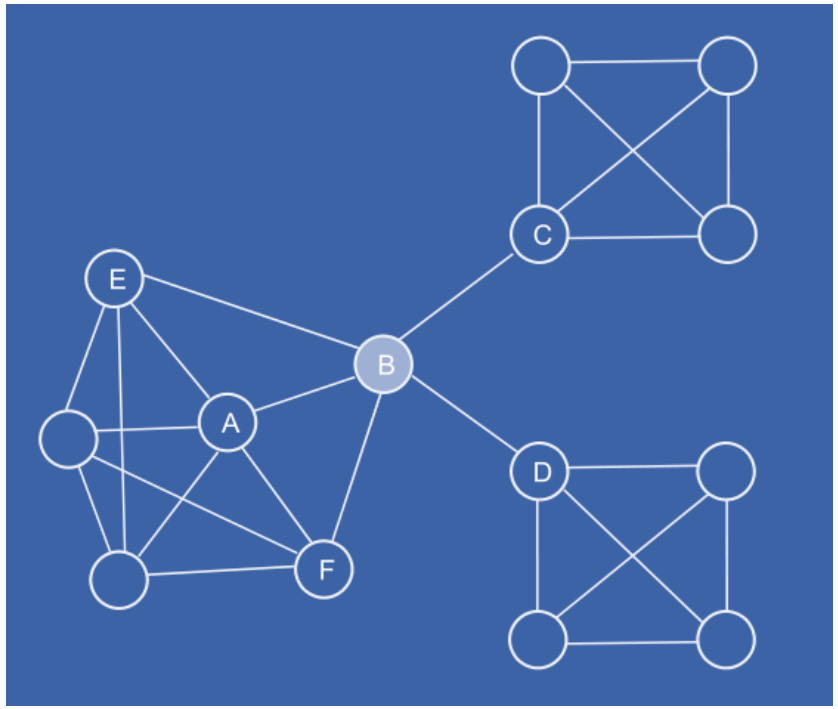
- **从三元闭包说起**

- 人们不是生来就结为好友的：新的好友关系从何而来？
  - 以三节点关系为例，如右图所示，受“三元闭包”的启发，当AB、AC关系已存在的时候，BC有较高概率也存在关系
    - 由此可以看出，部分网络中新增的好友关系是一种“被动”的增加，是受已有关系影响后形成的
      - 例如，通过已有好友的引荐，如“相亲”



- **关系的嵌入性衡量**

- 如果已有的两条边关系强度越强，最终形成三元闭包的可能性就越大
  - 关系强度的一种启发式衡量方法：嵌入性，即一条边两端节点**共同的好友数**
    - 例如右图中AB这一条边的嵌入性为2，因为有E、F两个共同好友
    - 嵌入性越高，这条边两端的节点之间的联系越紧密



- **最基本的链接预测**

- 基于嵌入性的启发，研究者们提出了最基本的链接预测方法
  - 两个节点之间存在一条边的概率，与它们共同好友的个数呈正比

$$s_{xy}^{CN} = |\Gamma(x) \cap \Gamma(y)|.$$

表示节点 y 的邻居集合

  - 该方法的合理性显而易见
    - 共同好友越多，两个人潜在的交集越大，被介绍认识的机会越多

- **最基本的链接预测：改进版**

- 然而，单纯基于共同邻居个数，会导致好友越多，越容易获得新好友
  - 如何削弱好友个数的影响？对该指标进行Normalization，例如：
    - Salton Index $$s_{xy}^{Salton} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{k_x \times k_y}},$$
    - Jaccard Index $$s_{xy}^{Jaccard} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}.$$
    - Hub Promoted Index $$s_{xy}^{HPI} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\min\{k_x, k_y\}}.$$
    - Hub Depressed Index $$s_{xy}^{HDI} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\max\{k_x, k_y\}}.$$

- **链接预测：另一种思路**

- 与此同时，部分指标仍坚持无尺度（Scale-free）网络的原则
  - 即，各节点的度数之间存在较大差异，且符合幂律（Power Law）分布
  - 在此种情况下，已有更多好友的节点更容易获得新好友
    - Preferential Attachment Index $s_{xy}^{PA} = k_x \times k_y,$
    - 事实上，无尺度网络确实在很多场景下符合实际
      - 因此，PA Index也具有一定的合理性
      - 此外，该指标的好处之一在于无需网络结构信息和邻居信息

- **链接预测：另一种思路**

- 与PA Index类似的思路
  - 即，如果某个节点已有的边越稠密，则这个节点越容易获得更多的边
  - 采用聚类系数以衡量节点已有边的稠密性

$$\textbf{clustering coef.}(v) = \frac{3 \times \#\text{ triangles adjacent to } u}{\#\text{ possible triples adjacent to } u}$$

  - 当需要判断(u, v)边存在的概率，和PA Index一样，将两个聚类系数乘起来

- **链接预测：邻居的价值**

- 先前的多数指标基于一个简单的假设：共同好友越多，越可能建立关系
  - 这一假设是建立在共同好友作为中介的引荐基础之上的
  - 然而，共同好友的引荐力度如何?
    - 与PageRank类似，节点的出边越多，权威性加成被平摊得越稀薄
    - 因此，不仅要考虑共同好友数量，还要考虑好友的好友数
    - Resource Allocation Index

$$s_{xy}^{RA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z}.$$

- **链接预测：从局部到全局**

- 先前的多数指标基于一个简单的假设：共同好友越多，越可能建立关系
  - 然而，建立好友关系这件事，出力的或许不只是共同好友
  - 例如，相亲这件事儿，可能先由父母联系了“媒婆”，“媒婆”作为中介进行介绍
    - 然而在这一场景中，“媒婆”并不是男女双方的共同好友

- **链接预测：从局部到全局**

- 不仅考虑“共同好友”，还考虑“共同好友”的好友
  - 基于多跳关系的链接预测基本方法
    - Katz Index $$s_{xy}^{Katz} = \sum_{l=1}^{\infty} \beta^l \cdot |paths_{xy}^{<l>}| = \beta A_{xy} + \beta^2 (A^2)_{xy} + \beta^3 (A^3)_{xy} + \cdots,$$
    - 其中，A^n表示沟通x、y节点的，路径长度为 n 的路径数量
      - N = 1时，A即为共同好友数
    - 同时，β 为调节路径权重的参数
      - 在 β 较小的情况下，Katz指标其实很接近于共同邻居指标

- **链接预测：从局部到全局**

- 更为一般化的基于全局网络结构的链接预测方法
  - Random Walk with Restart，受PageRank启发的链接预测方法
    - 参考PageRank的基本公式：

$$\mathrm{PR}(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{\mathrm{PR}(p_j)}{L(p_j)}$$

    - 其中的(1-d)/N的部分，相当于以一定概率重新选择起点，即Restart

- **链接预测：从局部到全局**

- 更为一般化的基于全局网络结构的链接预测方法
  - Random Walk with Restart，受PageRank启发的链接预测方法
    - RWR思路与之类似，公式如下：

$$\vec{q_x} = cP^T\vec{q_x} + (1-c)\vec{e_x},$$

    - 其中，$\vec{e_x}$ 向量仅x节点对应的维度为1，其他为0
      - 换言之，相当于指定起点

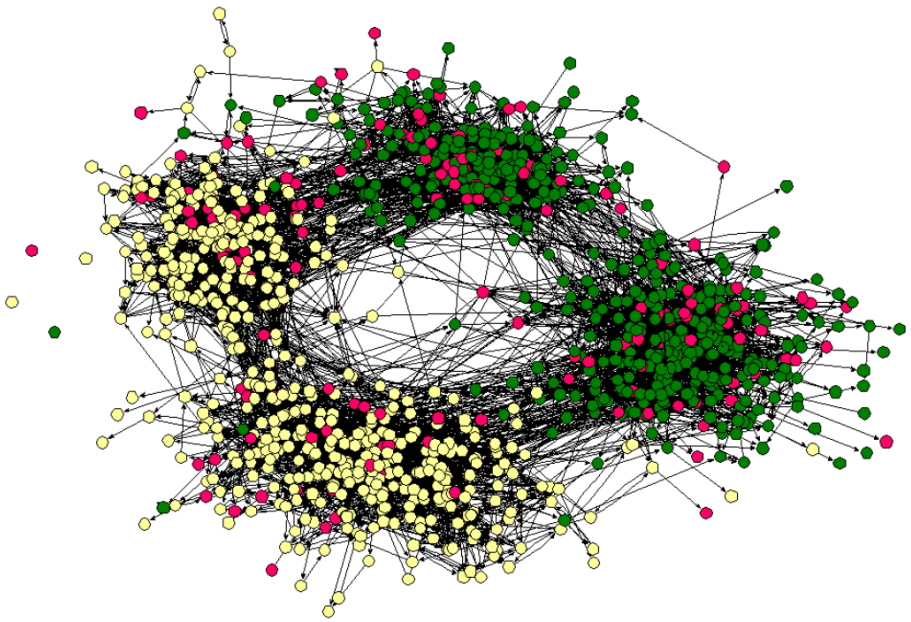
- **链接预测：从局部到全局**

- 更为一般化的基于全局网络结构的链接预测方法
  - Random Walk with Restart，受PageRank启发的链接预测方法
    - 显然，两个节点形成一条链接，是双方共同作用的结果。
    - 因此，最终的RWR指标计算公式为：

$$s_{xy}^{RWR} = q_{xy} + q_{yx}.$$

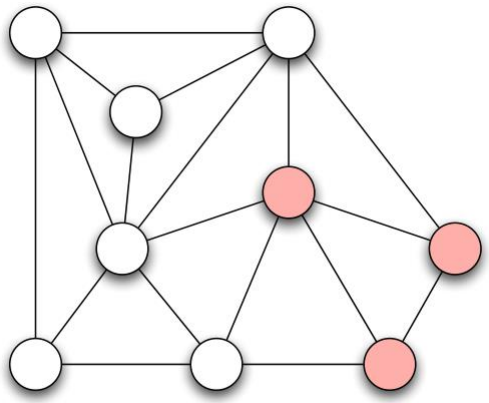
- **链接预测：属性的影响**

- 除了单纯考虑基于邻居关系的结构性影响外，属性也是决定链接的重要因素
  - 同质性的影响：人们在选择好友时，往往倾向于选择同类属性的节点
    - 例如，右图为某个镇中学的学生社会网络
    - 其中，从左到右的颜色分别代表不同的种族
    - 显然，图中呈现出明显的种族分化



• 链接预测：属性的影响

- 除了单纯考虑基于邻居关系的结构性影响外，属性也是决定链接的重要因素
  - 如何判断是否存在同质性因素的影响？
    - 考虑最简单的二分属性影响，比如性别因素
    - 如右图所示，假如我们有某个社会网络，其中 $p\%$为男性，$q\%$为女性，那么一条跨越性别的边的概率应为 $2pq$
    - 换言之，如果跨性别的边所占的比例远远小于$2pq$，则该社会网络体现出同质性影响的迹象



- **链接预测：属性的影响**

- 基于同质性的影响，也可以实现链接的预测
  - 例如，基于两个节点之间共享的属性来计算

$$\sum_{z\,:\,\text{feature shared by } x,y} \frac{1}{\log(frequency(z))}$$

  - 或者更为直接，将节点表征为向量，然后采用向量之间的相似性来衡量
    - 事实上，目前大多数基于网络表征实现链接预测的技术都沿用此思路
    - 有关网络表征的相关内容，可参考连德富老师的《深度学习》课程

- **链接预测：属性的影响**

- 然而，相似性并不是用户建立社交关系的充要条件
- 如何揭示节点在建立关系时存在的个人偏好?
  - 例如，能否根据已有的关系，判断节点更倾向于和哪些节点建立关系?
  - 思路：Random Walk （全局网络结构） + Pairwise LTR （训练偏好）
  - Supervised Random Walk（WSDM 2011论文）

论文作者：Jure Leskovec

Supervised random walks: predicting and recommending links in social networks

- **链接预测：属性的影响**

- 基于已有关系的有监督学习框架
  - Supervised Random Walk的两个模块
    - 基于偏序排序的目标函数

$$\min_{w} F(w) = ||w||^2$$

such that

$$\forall\, d \in D, l \in L: \; p_l < p_d$$

    - 随机游走部分 $Q_{uv} = (1-\alpha)Q'_{uv} + \alpha \mathbf{1}(v = s).$ $p^T = p^T Q$
      - 其最终目标是训练参数 w， 用于描述建立关系时的偏好选择

链接预测：其它的影响因素

- 除了前述的种种基本因素外，链接的生成还受到诸多因素的影响
  - 时间因素：喜新厌旧是人的本性，不同时间的链接有着不同参考价值
  - 范围约束：与全局性随机不同，有些节点注定不会相遇
    - 例如，基于位置的网络中，活动范围以外的节点不可能结为好友
  - 瞬时网络：受特定事件驱动，仅存在于特定时段内，之后会消亡
    - 瞬时网络与长期网络存在着相互转化的作用
  - 网络的符号性……

本节目录

- 网络生成模型
  - 随机网络
  - 幂律特性与B-A模型
- 链接预测模型
  - 基于邻居关系的模型
  - 同质性与链接选择
- **网络符号性**

先前回顾

- **强弱关系的区分**

- 上节课中，我们介绍了“三元闭包”与强弱关系
  - 虽然强弱关系功能不同，但对于个体来说，都是促进作用
- 然而，现实中的关系要远比强弱之分更复杂
  - 一类典型的情况是，网络中还存在着“朋友”（+）与“敌人”（-）之分
    - 例如，“五排”时的队友和对手
    - 此时，仅靠关系强弱、亲疏已经无法描述其关系

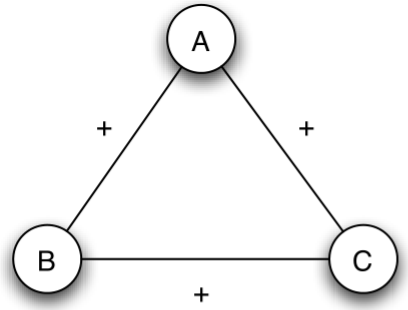
三角关系中的符号与平衡

- 我们先从最基本的三角关系入手
  - 因为三角形具有稳定性（大误）
- 从社会心理学角度，三角关系在以下两种情况时可以维持平衡

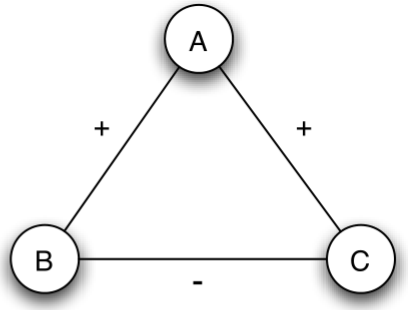
① 三条关系均为正（+）

② 三条关系为两负一正

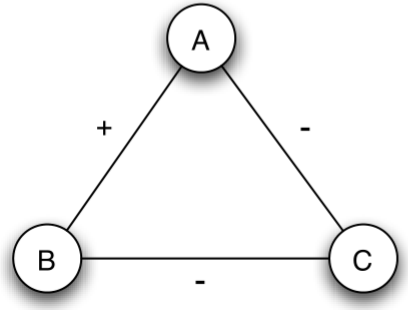
➢ 其中暗含的规律：负边个数为偶数



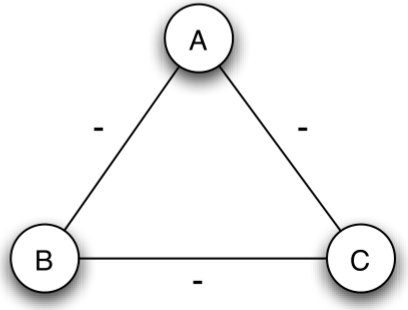
(a) *A, B, and C are mutual friends: balanced.*



(b) *A is friends with B and C, but they don't get along with each other: not balanced.*



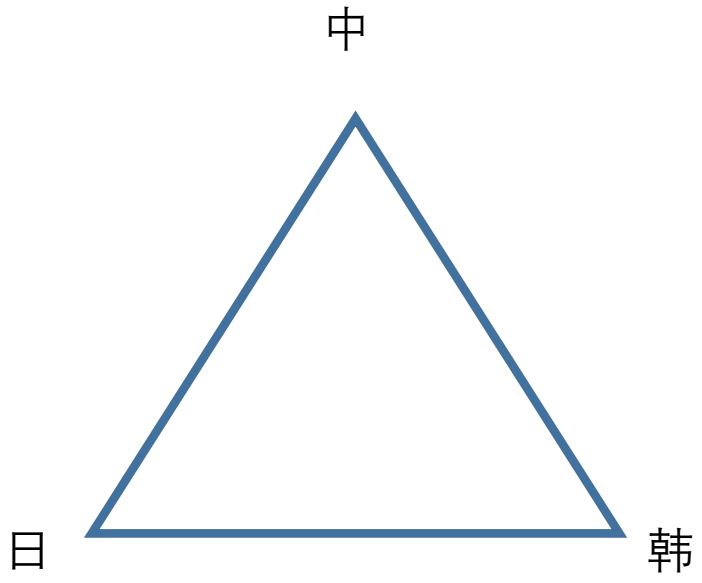
(c) *A and B are friends with C as a mutual enemy: balanced.*



(d) *A, B, and C are mutual enemies: not balanced.*

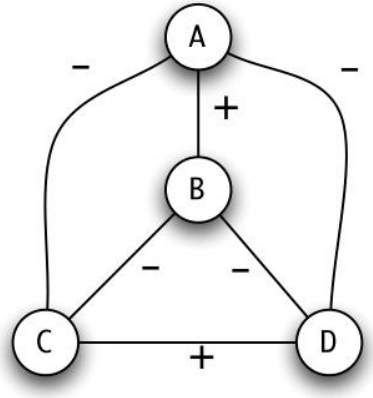
三角关系中的符号与平衡

- 我们先从最基本的三角关系入手
  - 上述规律对应着民间的俗语
    - “朋友的朋友是朋友，敌人的敌人是朋友”
- 实例：东亚三国外交
  - 坊间传言：“中日友好靠韩国，中韩友好靠日本”
    - 如何维持平衡?
    - 一点小变数：*外部作用*

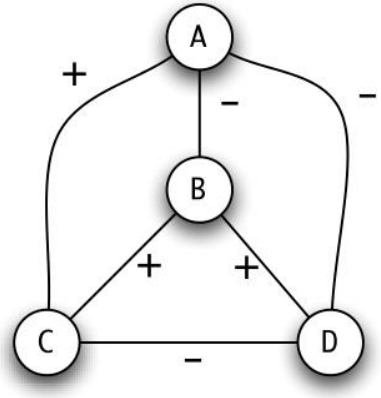


从三角关系到多点平衡

- 对于多个节点构成的完成图，其平衡性需要根据其中每一个三角形来判定
  - 平衡性的判定规则与三角关系的平衡性判定一致
    - 每个三角形中的三条边，当均都为“+”或恰好只有一条为“+”时平衡



balanced



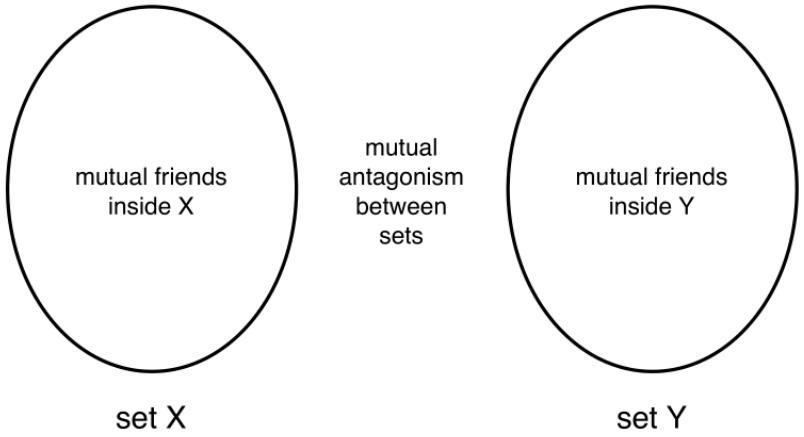
not balanced

- **从三角关系到多点平衡**

- 从多点平衡性判定中总结出的平衡定理
  - 如果一个标记的完全图是平衡的，那么（二选一）：

    ① 它的所有节点都是两两之间是朋友（正关系）

    ② 它的节点可以被分为两个组X和Y，其中X、Y两个组内的节点都是两两互为朋友，而X组内的每个节点都是Y组内每个节点的敌人



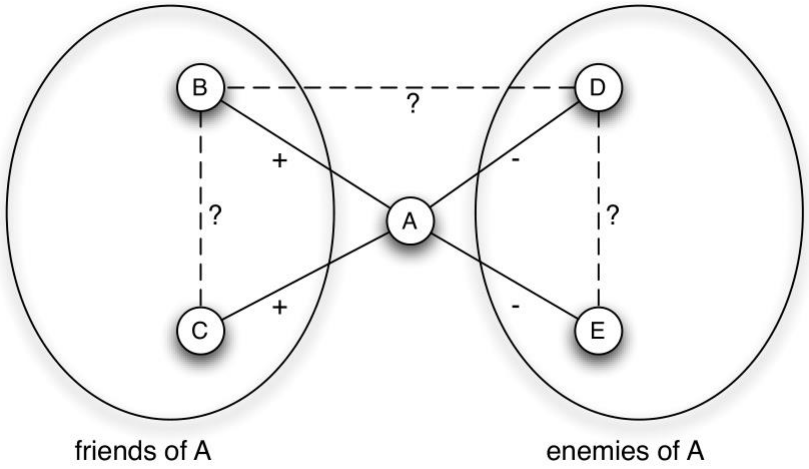
从三角关系到多点平衡

- 平衡定理的证明
  - 首先取任意一个节点 A，将其正关系和负关系的节点分别形成X、Y组合，那么：

    ① 对于X组合来说，假如任意两节点B、C不是朋友，则该三角形不平衡

    ② 对于Y组合来说，假如任意两节点D、E不是朋友，则该三角形不平衡

    ③ 对于跨组合的任意两节点B、D来说，假如它们是朋友，则该三角形不平衡



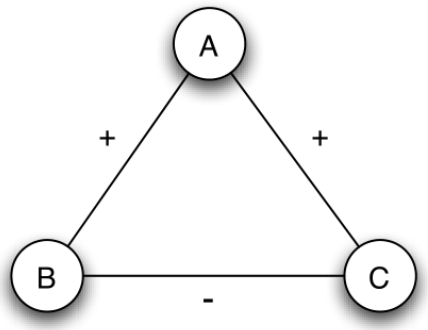
- **弱平衡现象**

- 在三角关系平衡性判定中，我们指出了两类不平衡的情况

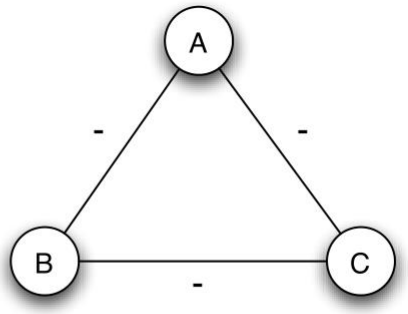
  ① 三者之间的关系为两正一负（+，+，-）

  ② 三者之间的关系均为负（-，-，-）

- 思考一下，这两种情况的动机和影响是一样的吗?



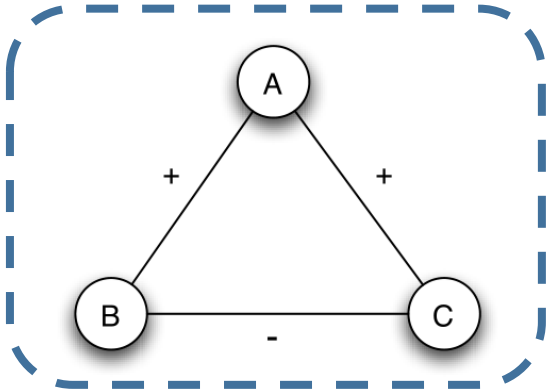
(b) *A is friends with B and C, but they don't get along with each other: not balanced.*



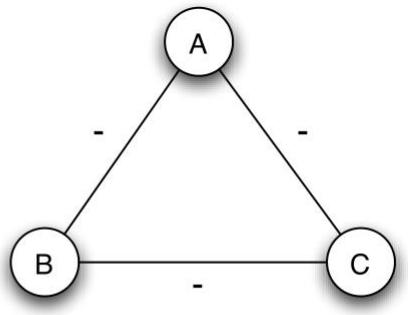
(d) *A, B, and C are mutual enemies: not balanced.*

弱平衡现象

- 从心理学角度来看，三者均为负的情况，要比两正一负更稳定
  - 理由：有共同好友时，共同好友会努力化解负面关系，但三者均为负时无此动力
- 由此，衍生出了所谓“弱平衡网络”
  - 定义：任意三个节点，均不存在两正一负（+，+，-）的连接模式



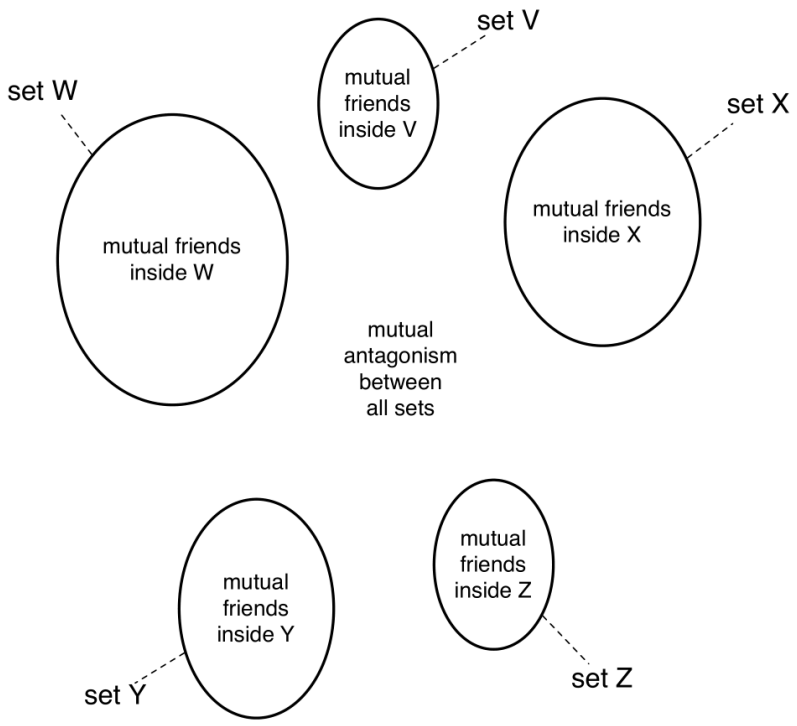
(b) *A is friends with B and C, but they don't get along with each other: not balanced.*



(d) *A, B, and C are mutual enemies: not balanced.*

弱平衡定理

- 结合弱平衡定义，衍生出的弱平衡定理
  - 如果一个标记的完全图是弱平衡的，那么：
    - ✓ 它的所有节点可以被分为多个不同的组
    - ✓ 每个组内任意两个节点互为朋友，不同组之间任意两个节点互为敌人



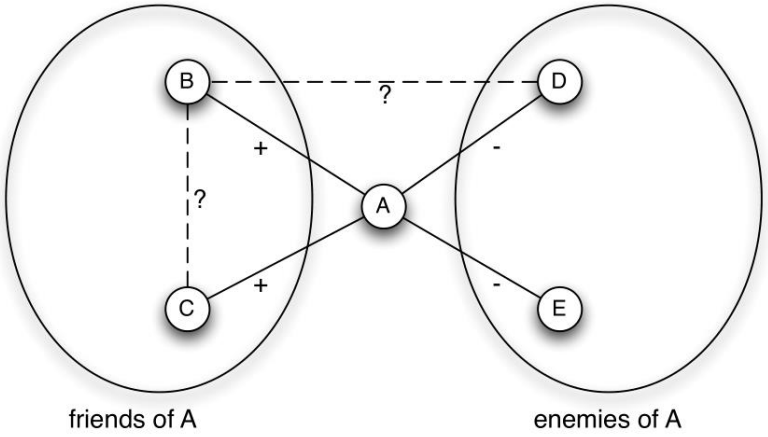
从三角关系到多点平衡

- 弱平衡定理的证明
  - 同样，取任意一个节点 A，将其正关系和负关系的节点分别形成X、Y组合，那么：

    ① 对于X组合来说，假如任意两节点B、C不是朋友，则该三角形不平衡

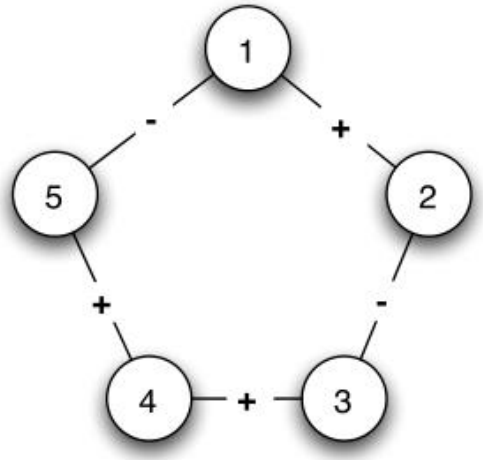
    ② 对于跨组合的任意两节点B、D来说，假如它们是朋友，则该三角形不平衡

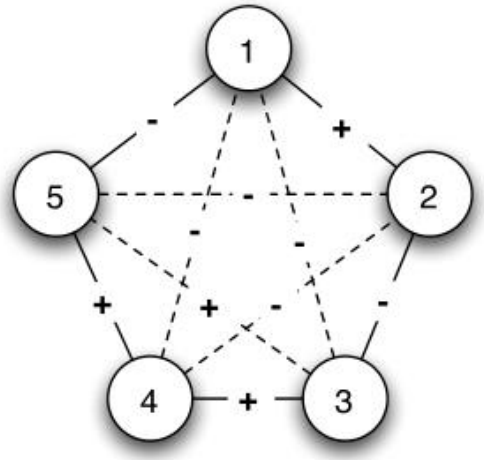
    ◆ 对比一下平衡定理，少了什么？



非完全网络的平衡问题

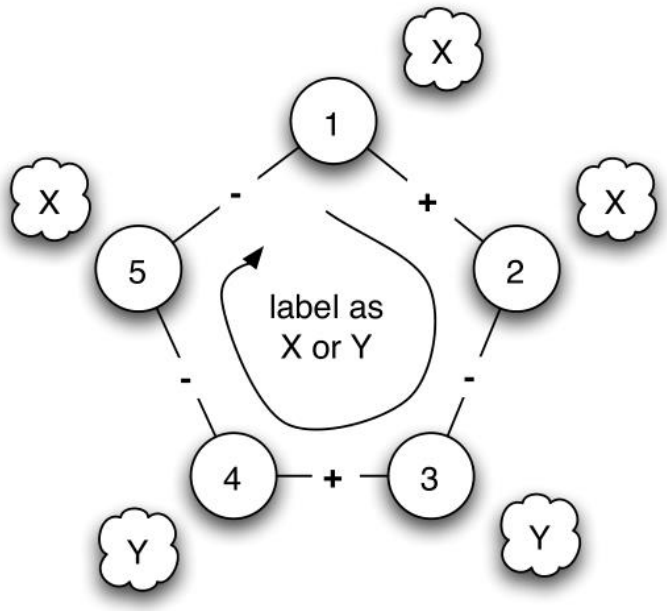
- 前面所讨论的平衡性，都是建立在完全图的基础之上
  - 对于非完全图的情况，由于三角形的不完整，如何判定其平衡性呢?
  - 启发式方法：基于符号平衡性，把非完全图填充为完全图
    - 如果可以找到填充方案，则该非完全网络是平衡的
    - 但是这种方法过于简单粗暴，想要完全填充也绝非易事





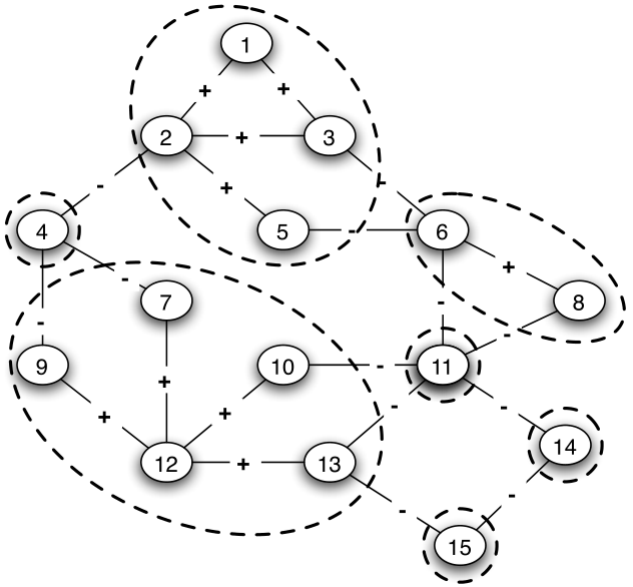
非完全网络的平衡问题

- 一种非完全网络平衡性的启发式判定方法
  - 在三角关系平衡性判定时，我们总结了一条规律：负边个数为偶数
  - 事实上，我们也发现，当负边个数为奇数时，我们无法将一个非完全网络划分为平衡性判定所需的X、Y两个组合，满足组内为好友，组间为敌人的约束
    - 划分方式：从 1 号节点开始分组，每遇到一个负关系就变换一次分组
    - 结果发现，当负关系个数为奇数时，最后一个节点无法进行分组



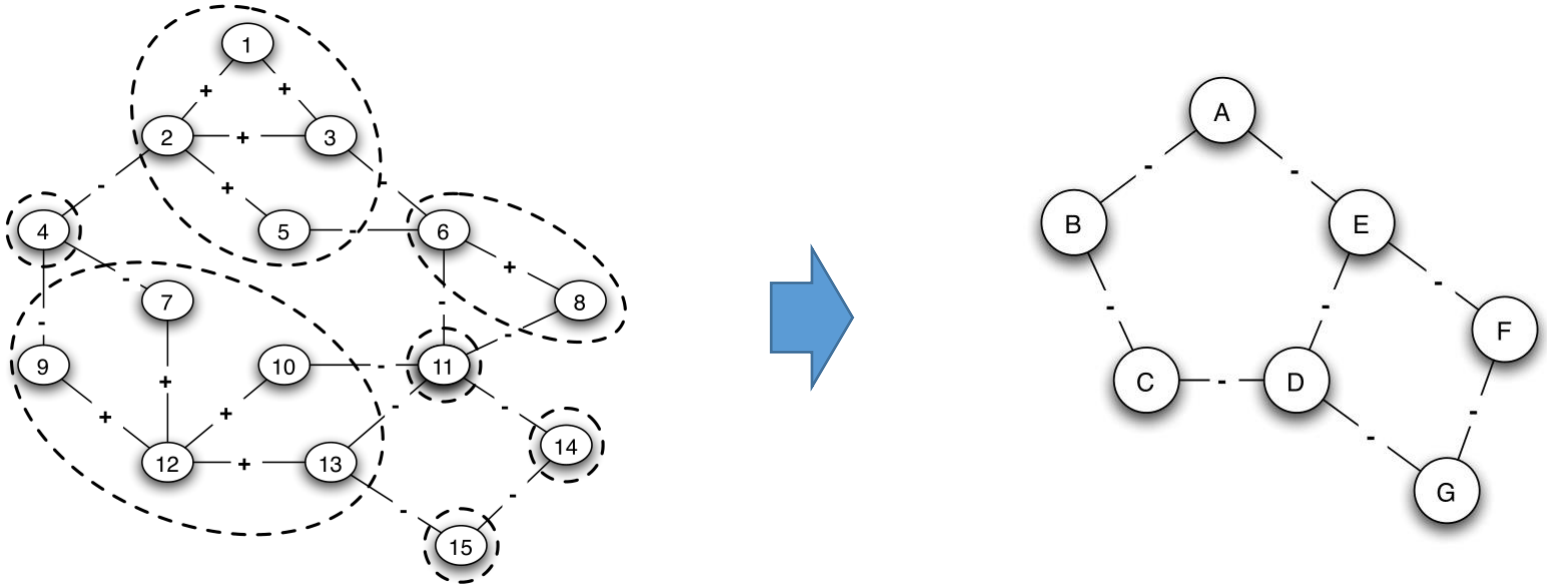
非完全网络的平衡问题

- 更加一般化的方式：抽取正连通子图
  - 前述启发式方法其实具有普遍适用性，但在复杂网络结构中很难直观看出
  - 我们还是从平衡性的判定定理出发
    - 平衡性的核心思想是分组，组内均为好友，组间均为敌人
    - 从图论角度考虑，每个分组即对应着一个正连通子图
  - 那么，能否把每一个正连通子图抽象成为一个节点?



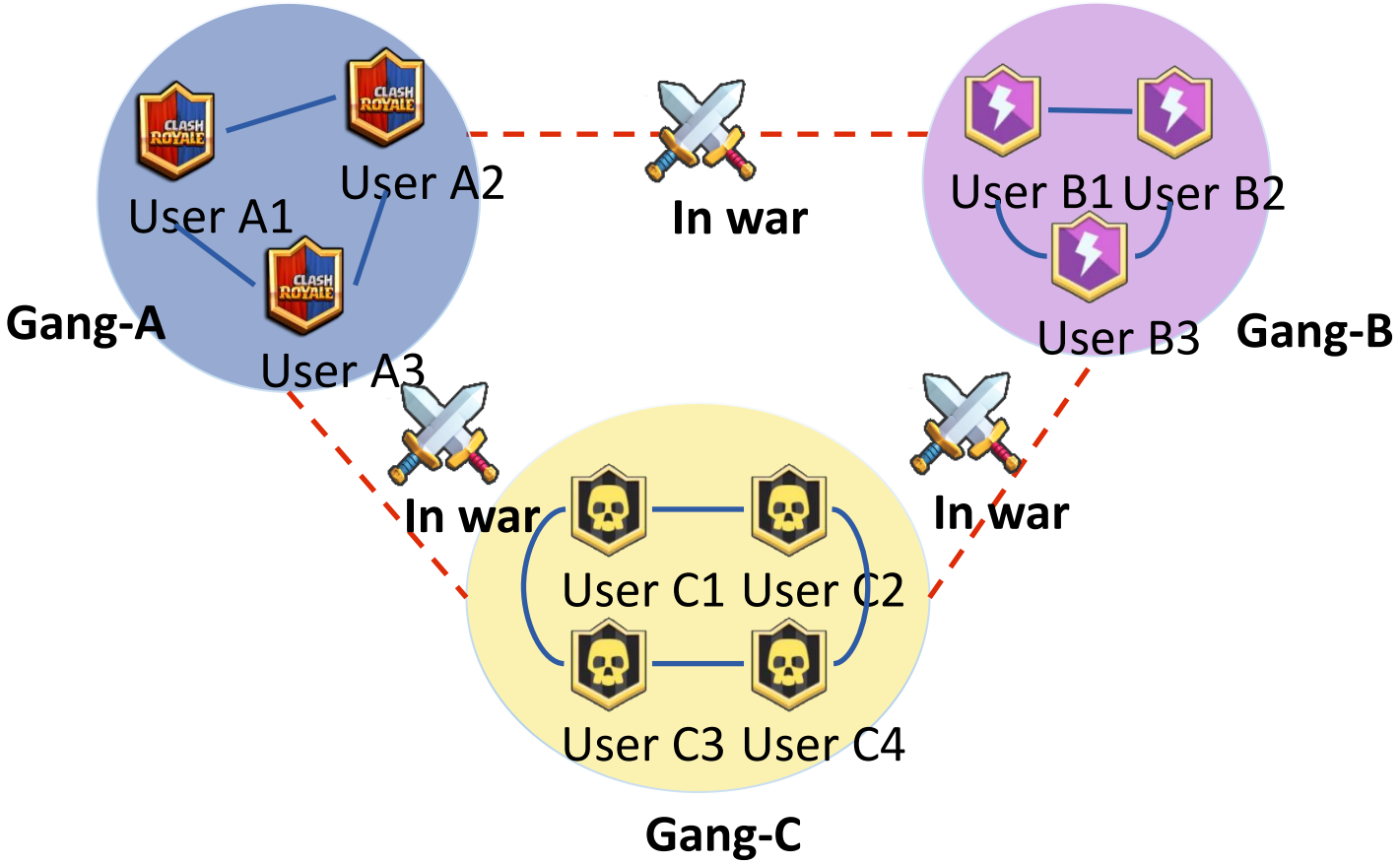
非完全网络的平衡问题

- 更加一般化的方式：抽取正连通子图
  - 通过寻找正连通子图并抽象为节点的方式，可以大大简化网络平衡性的判定难度
  - 同时通过这种“抽象”，也可以实现对于前述启发式方法的验证



- **更广泛的网络符号性问题**

- 更有趣的是，这种抽象的过程实际上可以揭示不同粒度主体间的合作/竞争关系
  - 现实网络中，不仅个体间存在友好/敌对，个体形成的团体同样有着类似的关系

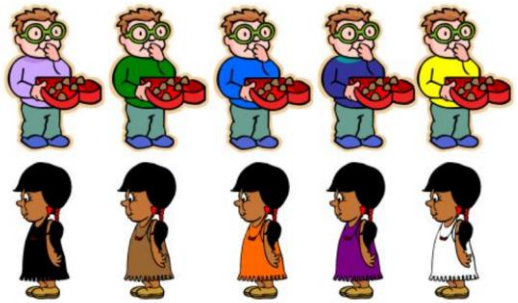


Lingyang Chu, et al., Finding Gangs in War from Signed Networks, KDD-2016

- **从网络符号性到博弈问题**

- 而当主体/团体间存在敌对/竞争，他们就必然需要通过寻求合作者来壮大自己的实力
  - 而在这个过程中，合作一方也将“待价而沽”，选择更合适/更有前途的一方加入
  - 一个该问题的简化版本，就是个体的婚姻配对问题

Stable Matching



- 如何将这种网络符号性拓展到竞争/合作，如何衡量这一过程中的收益，如何判断配对的稳定性？

——*详情将在匹配问题部分介绍*

- **尾声：符号的演化**

我们前面的讨论，都是围绕静态的符号网络展开

而在现实中，网络符号往往处于不断演化中

本章小结

网络链接与符号性

- **网络生成模型**
  - 随机网络
  - 幂律网络
- **链接预测模型**
- **网络符号性问题**

tongxu@ustc.edu.cn