

# 正关系与负关系

黄俊杰<sup>1</sup>

<sup>1</sup> 网络数据科学与技术重点实验室  
中国科学院计算研究所

2018 年 5 月 17 日

# Outline

Good morning

Warm up

Glossary

正负关系

结构平衡

结构平衡的应用

结构平衡的弱形式

结构平衡定义的推广

任意非完全网络的结构平衡

近似平衡的网络

习题

结构平衡推广

中国历史人物正负社交关系

数据描述

Signed Graph 构建

Inductive Representation

Learning on Large Graph

Introduction

Related Work

Method

Results

Ideas



# Warm up

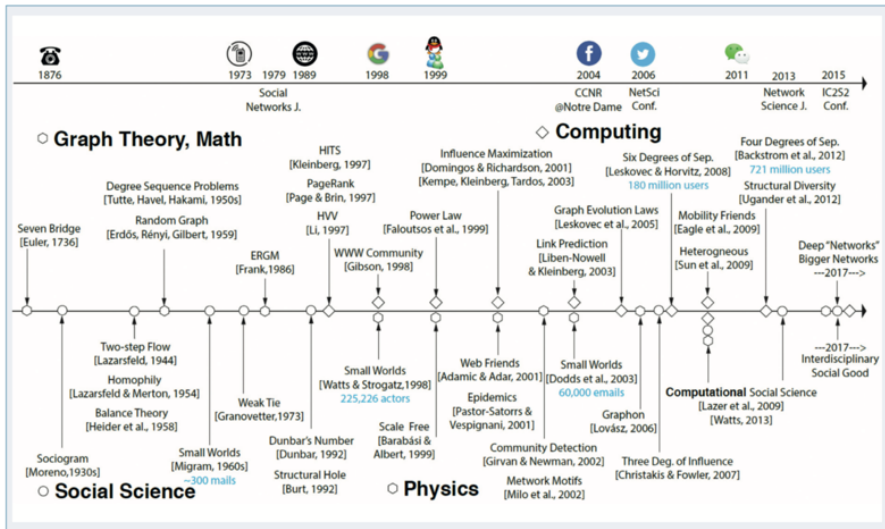


图1 社会网络科学简史 (该图源自微软研究院Yuxiao Dong博士)

# Glossary

- 邻居 neighbor
- 路径 path
- 圈 cycle
- 连通性 connectivity
- 连通分量 connected component
- 小世界现象 small world phenomenon
  - 聚类系数 the clustering coefficient
  - 度分布 degree distribution
  - 平均路径长度 average path length
- 三元闭包 triadic closure
- 桥 bridge & 捷径 local bridge
- 强关系弱关系 strong tie & weak tie
- 邻里重叠度 neighborhood overlap
- 结构洞 structural hole
- 图划分 graph partitioning
- 同质 homogeneous
- 异构 heterogeneous



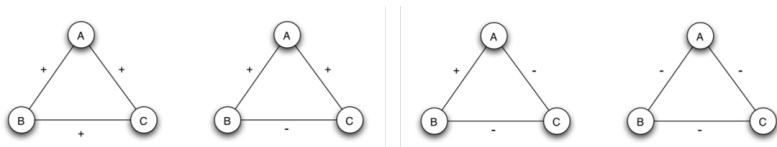
# 结构平衡的引入

- ▶ 将网络的边带上“正”或者“负”的涵义
- ▶ 结构平衡是理解两种力量之间的张力的基本框架
- ▶ 理解局部作用影响全局行为的方式
  - ▶ 所谓局部作用，即发生在少数几个节点上的现象
  - ▶ 所谓全局行为，即在整个网络层面能够观察到的现象
- ▶ Supporting Principle (proposed by Heider)  
my friend' s friend is my friend, the enemy of a friend is my enemy, my enemy' s enemy is my friend (or may become one), and so on



# 结构平衡

- ▶ 团 (clique) 或完全图 (complete graph)  
每对节点之间都有一条边
- ▶ 用 + 或-标示每条边，+ 标识两个端点是朋友，-标示两个端点之间是敌人
- ▶ 其关键思想是：对于一个三元组，有四种不同的组合

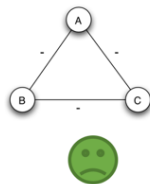
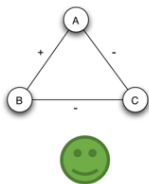
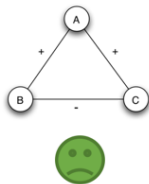
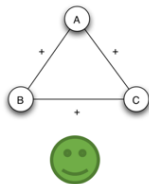


# 结构平衡

## 平衡性

对于一个完全图，其边以 + 或 - 标记，如果满足以下条件，则称其有平衡性

任意三个节点，要么都标示为 + 或者恰有一条边标示为 +



# 结构平衡网络的特性

## 平衡定理

如果一个标记的完全图是平衡的，则要么它的所有节点两两都是朋友，要么它的节点可以分成两个组 X 和 Y，其中 X 组内的节点两两都是朋友，Y 组内的节点两两也都是朋友，而 X 组中的每个节点都是 Y 组中每个节点的敌人。

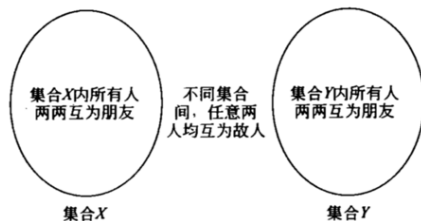


图 5.3 一个完全图是平衡的：如果该完全图可以划分成两个集合，每个集合内任意两个人均互为朋友，属于不同集合的任意两个人均互为敌人





# 平衡定理的证明

取网络中的任意一个节点，称为 A。

对 A 而言，其他节点要么是 A 的朋友，要么是 A 的敌人。

将朋友和 A 确定为 X, Y 则是包含 A 的所有敌人

## 需要证明的条件

1. X 中的每两个节点都是朋友
2. Y 中的每两个节点都是朋友
3. X 中的每个节点与 Y 中的每个节点都是敌人

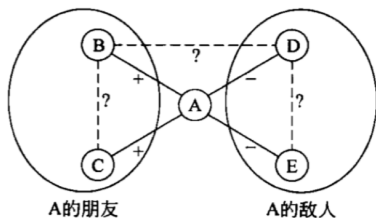


图 5.4 分析平衡网络的示意图(可能还包含其他没有画出的节点)



# 结构平衡的应用

## 1. 国际关系

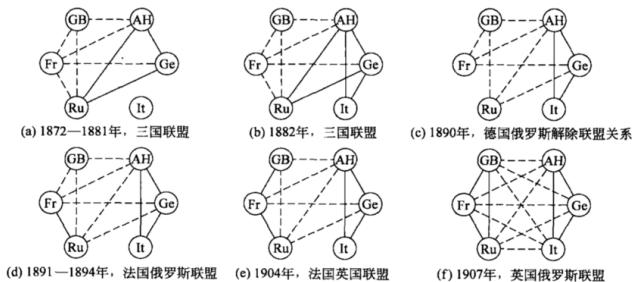


图 5.5 1872 年至 1907 年间，欧洲联盟的演化

## 2. 信任、怀疑和在线排名

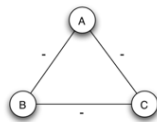
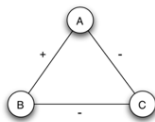
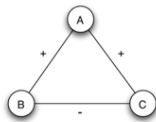
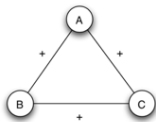


# 结构平衡的弱形式

## 弱平衡性

对于一个完全图，其边以 + 或 - 标记，如果满足以下条件，则称其有弱平衡性

任意三个节点，均不存在两个正关系边和一个负关系边的连接模式



# 弱结构平衡

## 弱平衡网络的特性

如果一个标记的完全图是弱平衡的，则要么它的所有节点可分成不同的组，并且满足同一组中的任意两个节点互为朋友。

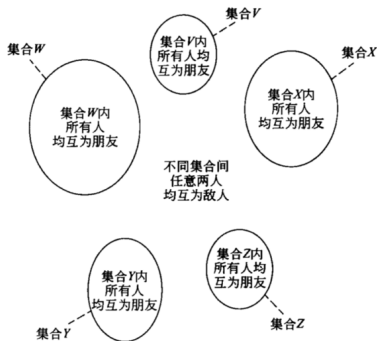


图 5.6 一个完全图满足弱平衡：该图可以分成多个组，每个组内任意两个人均互为朋友关系，不同组之间任意两个人均互为敌对关系



## 弱平衡特性的证明

取网络中的任意一个节点，称为 A。

设包含 A 及其所有朋友的集合为 X。设定两个条件

1. A 的所有朋友均互为朋友
2. A 及其朋友和图中除他们以外的所有人均互为敌人

在除去及其所有朋友的集合 X，定义为第一组，剩下的节点形成一个更小的弱平衡性的完全图，以此类推，直到所有节点都被分到各自的组中。

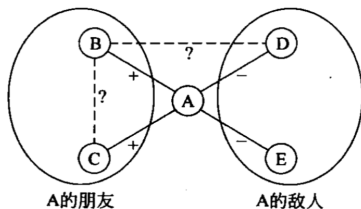


图 5.7 弱平衡网络的图示

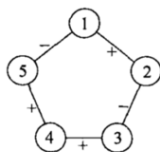
# 结构平衡定义的推广

- ▶ 非完全图的情况
- ▶ 放松为大多数三角形为平衡的

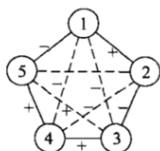


# 任意非完全网络的结构平衡

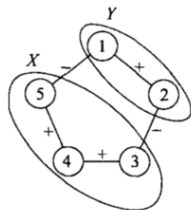
- ▶ 填充法：补全缺少的标示边，使其完全图平衡
- ▶ 全局划分默认法：默认划分为两个集合，集合内若相识则是正关系，集合间若相识为负关系



(a) 一个不完全图



(b) 填补缺失的边，形成一个完全的结构平衡图



(c) 将节点分成集合X和Y，使得集合内所有边为正，集合之间所有边为负

图 5.9 通过两种方法定义不完全图的结构平衡



# 近似平衡的网络

## 原始的平衡定理

若完全标注图中每个三角形都是平衡的，则：

每对节点都互为朋友

可将节点划分为两组集合 X 和 Y，满足：

- ▶ 集合 X 中每对节点都互为朋友
- ▶ 集合 Y 中每对节点都互为朋友
- ▶ 集合 X 中每个节点都与 Y 中每个节点互为敌人

## 放宽的平衡定理

若一个完全标注图中至少 99.9% 的三角形都是平衡的，则：

1. 存在一个至少包含 90% 节点的集合，由这些节点组成的对，其中至少 90% 的节点对互为朋友
2. 可将节点划分为两组集合 X 和 Y，满足：
  - ▶ 集合 X 中至少 90% 每对节点都互为朋友
  - ▶ 集合 Y 中至少 90% 每对节点都互为朋友
  - ▶ 分别属于集合 X 和 Y 的节点对至少有 90% 互为敌人





# 近似平衡的网络

## 放宽的平衡定理

若一个完全标注图中至少 99.9% 的三角形都是平衡的，则：

1. 存在一个至少包含 90% 节点的集合，由这些节点组成的对，其中至少 90% 的节点对互为朋友
2. 可将节点划分为两组集合 X 和 Y，满足：
  - ▶ 集合 X 中至少 90% 每对节点都互为朋友
  - ▶ 集合 Y 中至少 90% 每对节点都互为朋友
  - ▶ 分别属于集合 X 和 Y 的节点对至少有 90% 互为敌人

## 一般的平衡定理

设定  $\epsilon$  为任意数，范围为  $0 \leq \epsilon \leq \frac{1}{8}$ ，定义  $\delta = \sqrt[3]{\epsilon}$ 。若一个完全标注图中至少  $1 - \epsilon$  的三角形都是平衡的，则：

1. 至少存在  $1 - \epsilon$  占比的节点的，其中至少  $1 - \delta$  占比的节点对都是互为朋友
2. 可将节点划分为两组集合 X 和 Y，满足：
  - ▶ 集合 X 中至少  $1 - \epsilon$  占比的节点对都互为朋友
  - ▶ 集合 Y 中至少  $1 - \epsilon$  占比的节点对都互为朋友
  - ▶ 在集合 X 和 Y 之间的所有节点对中，至少有  $1 - \delta$  占比的节点对互为敌人



# 证明

对于一个完全图，假设一个图有  $N$  节点

## 1. 边和三角形的个数

- ▶ 边的总数为:  $C_n^2 = \frac{N(N-1)}{2}$
- ▶ 三角形的总数:  $C_n^3 = \frac{N(N-1)(N-2)}{6}$



# 证明

对于一个完全图，假设一个图有  $N$  节点

1. 边和三角形的个数
2. 找到一个良好节点

- ▶ 最不平衡三角形总数最多:  $\frac{\epsilon N(N-1)(N-2)}{6}$
- ▶ 设定一个节点的权重为它所涉及的不平衡三角形数，则
- ▶ 所有节点的权重之和:  $\frac{3\epsilon N(N-1)(N-2)}{6}$
- ▶ 一个节点的平均权重:  $\frac{\epsilon N(N-1)}{2}$ ，注意  $\frac{\epsilon N(N-1)}{2} < \frac{\epsilon N^2}{2}$
- ▶ 选泽权重等于或者低于平均值的节点  $A$  为良好节点，则最多在  $\frac{\epsilon N^2}{2}$  个三角形中



# 证明

对于一个完全图，假设一个图有  $N$  节点

1. 边和三角形的个数
2. 找到一个良好节点
3. 参照“良好节点”拆分图

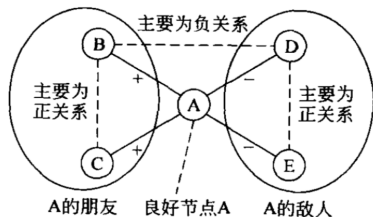


图 5.17 近似平衡完全图的特性遵循与平衡定理相似的论证分析方法

- ▶ X 和 Y 最多有  $\frac{\epsilon N^2}{2}$  条负边
- ▶ X 和 Y 集合之间的边构成的不平衡三角形最多有  $\frac{\epsilon N^2}{2}$  条正边
- ▶ 假设  $N$  是偶数,  $\delta N$  是整数, 且  $x$  为 X 的节点数,  $y$  为 Y 的节点数



# 证明

## · 的条件

1. 集合  $X$  中至少  $1 - \epsilon$  占比的节点对都互为朋友

假设  $x \geq (1 - \delta)N$ , 由  $\epsilon < \frac{1}{8}, \delta = \sqrt[3]{\epsilon}$ , 则  $\delta < \frac{1}{2}, x > \frac{1}{2}N$

集合  $X$  中有  $x$  个节点, 所以有  $\frac{x(x-1)}{2}$  边, 由于  $x > \frac{1}{2}N$ , 边数至少是

$\frac{(\frac{1}{2}N+1)(\frac{1}{2}N)}{2} \geq \frac{(\frac{1}{2}N)^2}{2} = \frac{N^2}{8}$  则占比最多为

$$\frac{\epsilon N^2 / 2}{N^2 / 8} = 4\epsilon = 4\delta^3 < \delta$$



# 证明

## · 的条件

1. 集合  $X$  中至少  $1 - \epsilon$  占比的节点对都互为朋友
2. 集合  $Y$  中至少  $1 - \epsilon$  占比的节点对都互为朋友

$X, Y$  之间有  $xy$  条边  
 $xy \geq \delta N(1 - \delta)N \geq \delta \frac{N^2}{2}$  (中文本这里出错了)

$$\frac{\epsilon N^2 / 2}{\delta N^2 / 2} = \frac{\epsilon}{\delta^2} = \delta^2 < \delta$$



# 证明

## · 的条件

1. 集合  $X$  中至少  $1 - \epsilon$  占比的节点对都互为朋友
2. 集合  $Y$  中至少  $1 - \epsilon$  占比的节点对都互为朋友
3. 在集合  $X$  和  $Y$  之间的所有节点对中, 至少有  $1 - \delta$  占比的节点对互为敌人

集合  $X$  中, 总共有  $\frac{x(x-1)}{2}$  条边, 由于  $x > \delta N$ , 则边的总数至少为  $(\delta N + 1)(\delta N)/2 \geq (\delta N)^2/2$ , 占比

$$\frac{\epsilon N^2/2}{(\delta N)^2/2} = \frac{\epsilon}{\delta^2} = \delta$$



## 习题 1

假设一组人类学家正在研究三个互为邻里的小村庄组成的集合。每个村庄都有 30 人，包括 2 ~ 3 个大家庭。每个村庄的人们都互相了解自己村庄的人。

当人类学家在这三个村庄都建立一个社会网络，会发现人们都和自己村庄的人成为朋友，和其他两个村庄的人成为敌人。这就给出了 90 人形成的网络，该网络的边也会带有正关系和负关系的标识。

根据本章的定义，这个 90 人的形成的网络是平衡的吗？





## 习题 2

分析图中的网络关系：每对节点间都有一条边，5 条边形成正关系，另 5 条边形成负关系。网络中每条边都参与三个三角形：一条边与其他不在一条边上的节点形成了三角形。如由 A-B 边组成的三角形有 ABC、ABD、ABE。可以用同样的方法列出每条边参与的三角形。

对于每条边来说，它们参与的三角形中有多少是平衡的？又有多少是不平衡的？注意：由于网络的对称性，每条正关系边和每条负关系边数量应该是一样的，因此应该将这两种边的属性考虑进行。

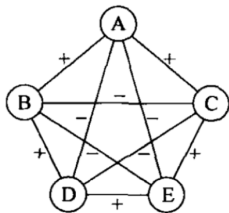


图 5.18 由 5 条正关系和负关系的边组成的网络

## 习题 3

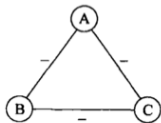


图 5.21 在由三个节点组成的社会网络中,每对节点均互为敌人

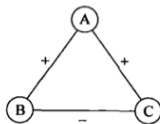


图 5.22 节点 A 与 B、C 为朋友,而 B 与 C 互为敌人

如何将第四个节点 D 加入到网络中,不会产生不平衡的三角形



## 习题 4

你和人类学家一起研究一个人烟稀少的热带雨林，其中 50 个农民生活在一条长 50 英里的河流沿岸。每个农民居住并占有沿岸的土地 1 英里长，因此他们完全瓜分了 50 英里长的河岸。

农民都相互认识，和他们交谈后，发现每个农民与其他住在 20 英里内的农民都是朋友，与其他 20 英里外的农民都是敌人。构建一个标注的完全图，对应这种社会关系，分析它是否具有结构平衡的特性，解释你的答案。



# Outline

Good morning

Warm up

Glossary

正负关系

结构平衡

结构平衡的应用

结构平衡的弱形式

结构平衡定义的推广

任意非完全网络的结构平衡

近似平衡的网络

习题

结构平衡推广

中国历史人物正负社交关系

数据描述

Signed Graph 构建

Inductive Representation

Learning on Large Graph

Introduction

Related Work

Method

Results

Ideas

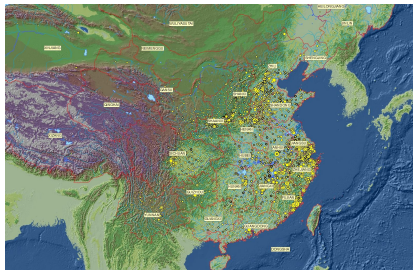


# CBDB 数据库

中国历代人物传记资料是线上的关系型资料库，目标在于系统性地收入中国历史上所有重要的传记资料，并将其内容毫无限制地、免费地公诸学术之用。

截至 2016 年 4 月为止，本资料库共收录约 370,000 人的传记资料，这些人物主要出自七世纪至十九世纪

本资料库除可作为人物传记的一种参考资料外，亦冀可敷统计分析与时空分析之用。



## 来源和现状

中国历代人物传记资料库之始祖为郝若贝教授（Robert M. Hartwell）（1932 -1996）

郝若贝教授将本资料库初版及其他财产遗赠哈佛燕京学社。

目前本资料库的开发工作係下属机构合作：

- ▶ 哈佛大學費正清中國研究中心
- ▶ 中央研究院歷史語言研究所
- ▶ 北京大學中國古代史研究中心



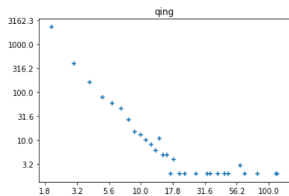
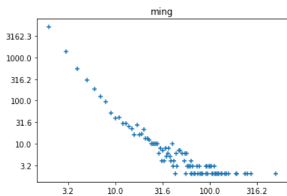
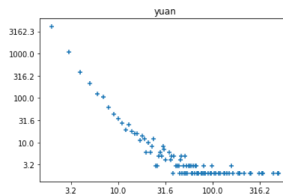
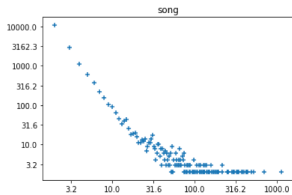
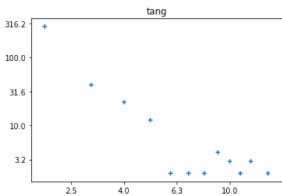
# 数据获取与预处理

- ▶ 来源于 CBDB API, 例子王安石
- ▶ 朝代划分
  - ▶ 标有的朝代
  - ▶ 出生年份在朝代年份内
  - ▶ 死亡年份在朝代年份内
- ▶ 统计结果

	$ \mathcal{V} $	$ \mathcal{E} $	Clustering Coefficient	Average Path Length
tang	365	286	0.016	1.6
song	17114	30330	0.121	4.08
yuan	6424	11864	0.150	4.001
ming	8350	14609	0.070	4.65
qing	3128	3059	0.021	7.71



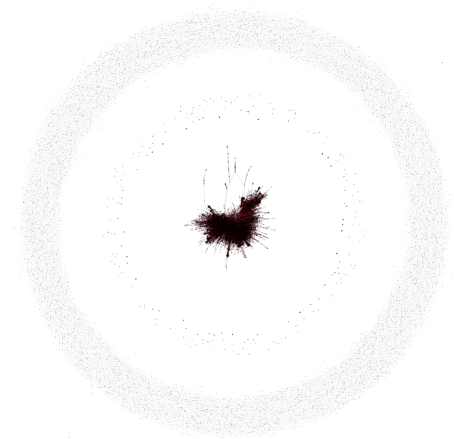
# 小世界网络





# 宋

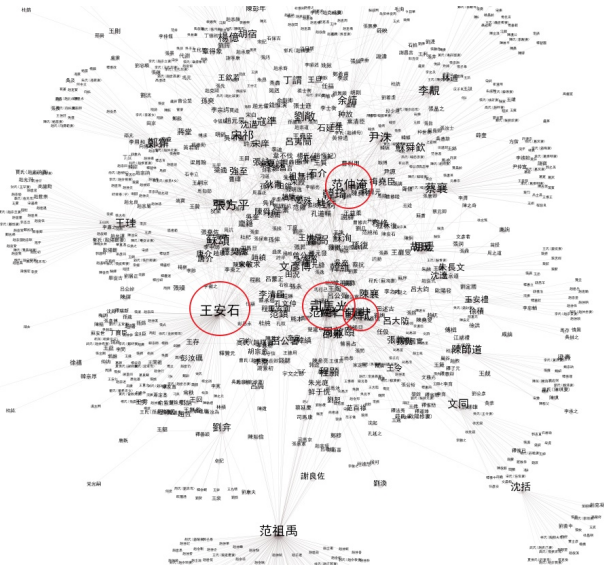
## ► 全宋人物关系



图：全宋人物关系



▶ 王安石 (1021-1086), 出生时间在 1000-1100 年的人物关系

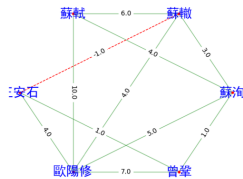
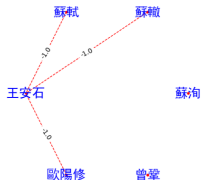
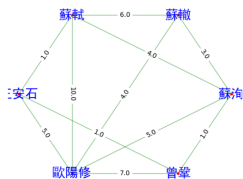


范祖禹

# Signed Graph 构建

手工对社交关系共计 445 种，进行标示为-1, 0, 1

- ▶ 所有朝代记录的统计结果为 positive(113029)
- ▶ 最多的正关系: 為 Y 作墓誌銘 (10189), 最多的负关系: 弹劾 (-400)
- ▶ 王安石和欧阳修政见不合, 但是关系为正
- ▶ 唐宋八大家关系图 (宋)

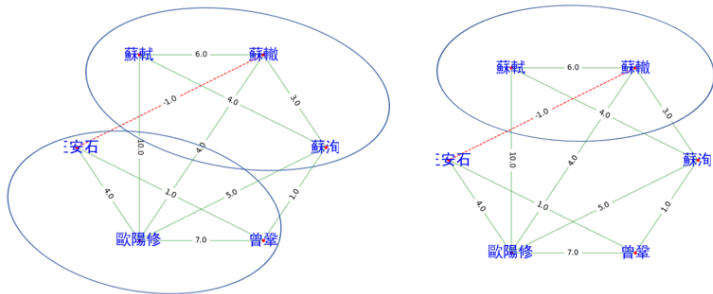


# 图划分与社区发现算法

$$\sum_{ij} \lambda A_{ij}^- \delta(\sigma_i, \sigma_j) + (1 - \lambda) A_{ij}^+ (1 - \delta(\sigma_i, \sigma_j))$$

$\delta(\sigma_i, \sigma_j)$  if  $\sigma_i = \sigma_j$  and 0 otherwise.  $\lambda$  type balance

$$\sum_{ij} X_{ij} = f_i$$



# Outline

Good morning

Warm up

Glossary

正负关系

结构平衡

结构平衡的应用

结构平衡的弱形式

结构平衡定义的推广

任意非完全网络的结构平衡

近似平衡的网络

习题

结构平衡推广

中国历史人物正负社交关系

数据描述

Signed Graph 构建

Inductive Representation

Learning on Large Graph

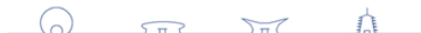
Introduction

Related Work

Method

Results

Ideas



# 简介

**Inductive representation learning** on large graphs. from NIPS2017.

Here we present **GraphSAGE**, a general inductive framework that leverage node feature information to efficiently generate **node embeddings** for each node.

▶ Author

1. William L. Hamilton
2. Rex Ying
3. Jure Leskovec



# 简介

**Inductive representation learning** on large graphs. from NIPS2017.

Here we present **GraphSAGE**, a general inductive framework that leverage node feature information to efficiently generate **node embeddings** for each node.

- ▶ Author
- ▶ Inductive VS Transductive

## Transductive Learning

转导学习是一种通过观察特定的训练样本，进而预测特定的测试样本的方法  
类比英美法系，实际案例直接结合过往的判例进行判决。

## Inductive Learning

归纳学习先从训练样本中学习得到通过的规则，再利用规则判断测试样本  
类比大陆法系，先对过往的判例归纳总结出法律条文，再应用到实际案例进行判决。



# 简介

**Inductive representation learning** on large graphs. from NIPS2017.

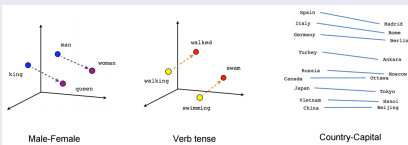
Here we present **GraphSAGE**, a general inductive framework that leverage node feature information to efficiently generate **node embeddings** for each node.

- ▶ Author
- ▶ Inductive VS Transductive
- ▶ Representation Learning node embeddings

## Representation learning

表示学习又称表征学习，其利用机器学习技术自动获取每一个实体或者关系的向量化表达。

Word2Vec 就是一套非常成功的表示学习技术，它将单词转化成了向量，并且可以让语义上相似的向量具有很高的相似度



Word2Vec 在一定程度上学习到了单词之间的抽象关系。  
如男人 - 女人 = 国王 - 王后





# 简介

**Inductive representation learning** on large graphs. from NIPS2017.

Here we present **GraphSAGE**, a general inductive framework that leverage node feature information to efficiently generate **node embeddings** for each node.

- ▶ Author
- ▶ Inductive VS Transductive
- ▶ Representation Learning node embeddings
- ▶ GraphSAGE: sampling and aggregation

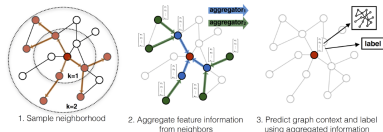


Figure 1: Visual illustration of the GraphSAGE sample and aggregate approach.



## 任务：节点分类 node classification

- ▶ 将学术论文分为不同的主题
- ▶ 将 Reddit 帖子分为不同的社区主题
- ▶ 将蛋白质交互网络用来区分蛋白质功能

	WoS	Reddit	PPI Tissue
$ \mathcal{V} $	302423	232965	
$ \mathcal{E} $	$9.15 \cdot 302423$	$232965 \cdot 492$	$35000 \cdot 4$
<i>Label</i>	6	50	2



## 相关工作

- ▶ Factorization-based embedding approach
  - ▶ Deepwalk
- ▶ Supervised learning over graph
- ▶ Graph convolutional networks



# Word2Vec

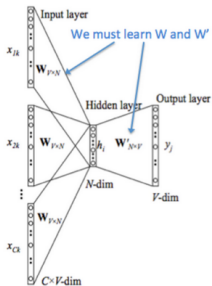


Figure 1: This image demonstrates how CBOW works and how we must learn the transfer matrices

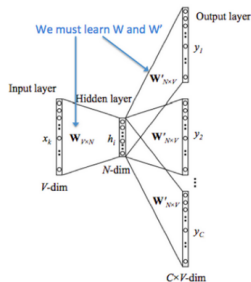


Figure 2: This image demonstrates how Skip-Gram works and how we must learn the transfer matrices

CBOW

$$-\log P(w_c | w_{c-m}, \dots, w_{c-1}, w_{c+1}, \dots, w_{c+m})$$

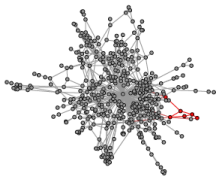
Skip-Gram

$$-\log P(w_{c-m}, \dots, w_{c-1}, w_{c+1}, \dots, w_{c+m} | w_c)$$

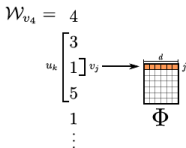
🖼️: Continous Bag of Words Model(CBOW) 和 Skip-Gram Model from cs224



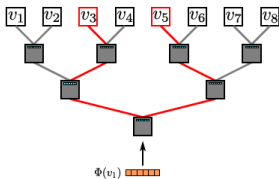
# Deepwalk



(a) Random walk generation.



(b) Representation mapping.



(c) Hierarchical Softmax.

---

## Algorithm 1 DEEPWALK( $G, w, d, \gamma, t$ )

---

**Input:** graph  $G(V, E)$

window size  $w$

embedding size  $d$

walks per vertex  $\gamma$

walk length  $t$

**Output:** matrix of vertex representations  $\Phi \in \mathbb{R}^{|V| \times d}$

1: Initialization: Sample  $\Phi$  from  $\mathcal{U}^{|V| \times d}$

2: Build a binary Tree  $T$  from  $V$

3: **for**  $i = 0$  to  $\gamma$  **do**

4:  $\mathcal{O} = \text{Shuffle}(V)$

5: **for each**  $v_i \in \mathcal{O}$  **do**

6:  $\mathcal{W}_{v_i} = \text{RandomWalk}(G, v_i, t)$

7: SkipGram( $\Phi, \mathcal{W}_{v_i}, w$ )

8: **end for**

9: **end for**

---

---

## Algorithm 2 SkipGram( $\Phi, \mathcal{W}_{v_i}, w$ )

---

1: **for each**  $v_j \in \mathcal{W}_{v_i}$  **do**

2: **for each**  $u_k \in \mathcal{W}_{v_i}[j - w : j + w]$  **do**

3:  $J(\Phi) = -\log \Pr(u_k | \Phi(v_j))$

4:  $\Phi = \Phi - \alpha * \frac{\partial J}{\partial \Phi}$

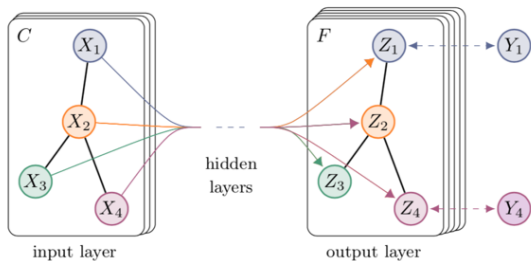
5: **end for**

6: **end for**

---



# GCN



(a) Graph Convolutional Network



(b) Hidden layer activations

图: GCN 算法示意

将非欧式距离的图结构，转换到傅立叶空间，然后进行卷积



# Embedding generation algorithm

---

**Algorithm 1:** GraphSAGE embedding generation (i.e., forward propagation) algorithm

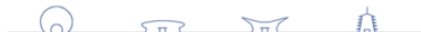
---

**Input** : Graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$ ; input features  $\{\mathbf{x}_v, \forall v \in \mathcal{V}\}$ ; depth  $K$ ; weight matrices  $\mathbf{W}^k, \forall k \in \{1, \dots, K\}$ ; non-linearity  $\sigma$ ; differentiable aggregator functions  $\text{AGGREGATE}_k, \forall k \in \{1, \dots, K\}$ ; neighborhood function  $\mathcal{N} : v \rightarrow 2^{\mathcal{V}}$

**Output** : Vector representations  $\mathbf{z}_v$  for all  $v \in \mathcal{V}$

```
1  $\mathbf{h}_v^0 \leftarrow \mathbf{x}_v, \forall v \in \mathcal{V}$  ;
2 for  $k = 1 \dots K$  do
3   for  $v \in \mathcal{V}$  do
4      $\mathbf{h}_{\mathcal{N}(v)}^k \leftarrow \text{AGGREGATE}_k(\{\mathbf{h}_u^{k-1}, \forall u \in \mathcal{N}(v)\})$ ;
5      $\mathbf{h}_v^k \leftarrow \sigma(\mathbf{W}^k \cdot \text{CONCAT}(\mathbf{h}_v^{k-1}, \mathbf{h}_{\mathcal{N}(v)}^k))$ 
6   end
7    $\mathbf{h}_v^k \leftarrow \mathbf{h}_v^k / \|\mathbf{h}_v^k\|_2, \forall v \in \mathcal{V}$ 
8 end
9  $\mathbf{z}_v \leftarrow \mathbf{h}_v^K, \forall v \in \mathcal{V}$ 
```

---



# Learning parameter and Loss function

- ▶ Supervised Loss

*crossEntropy loss*

- ▶ Unsupervised Loss

$$J_G(\mathbf{z}_u) = -\log\left(\sigma(\mathbf{z}_u^\top \mathbf{z}_v)\right) - Q \cdot \mathbb{E}_{v_n \sim P_n(v)} \log\left(\sigma(-\mathbf{z}_u^\top \mathbf{z}_{v_n})\right)$$

$v$  is a node that co-occurs near  $u$  on fixed-length random walk,  $\sigma$  is the sigmoid function,  $P_n$  is a negative sampling distribution, and  $Q$  defines the number of negative samples.





# Aggregator Architectures

- ▶ Mean Aggregator

$$\mathbf{h}_v^k \leftarrow \sigma(\mathbf{W} \cdot \text{MEAN}(\{\mathbf{h}_v^{k-1}\} \cup \{\mathbf{h}_u^{k-1}, \forall u \in \mathcal{N}(v)\}))$$

- ▶ LSTM Aggregator
- ▶ Pooling Aggregator

$$\text{AGGREGATE}_k^{\text{pool}} = \max(\{\sigma(\mathbf{W}_{\text{pool}} \mathbf{h}_{u_i}^k + \mathbf{b}), \forall u_i \in \mathcal{N}(v)\})$$



# Results

Table 1: Prediction results for the three datasets (micro-averaged F1 scores). Results for unsupervised and fully supervised GraphSAGE are shown. Analogous trends hold for macro-averaged scores.

Name	Citation		Reddit		PPI	
	Unsup. F1	Sup. F1	Unsup. F1	Sup. F1	Unsup. F1	Sup. F1
Random	0.206	0.206	0.043	0.042	0.396	0.396
Raw features	0.575	0.575	0.585	0.585	0.422	0.422
DeepWalk	0.565	0.565	0.324	0.324	—	—
DeepWalk + features	0.701	0.701	0.691	0.691	—	—
GraphSAGE-GCN	0.742	0.772	<b>0.908</b>	0.930	0.465	0.500
GraphSAGE-mean	0.778	0.820	0.897	0.950	0.486	0.598
GraphSAGE-LSTM	0.788	0.832	<b>0.907</b>	<b>0.954</b>	0.482	<b>0.612</b>
GraphSAGE-pool	<b>0.798</b>	<b>0.839</b>	0.892	0.948	<b>0.502</b>	0.600
% gain over feat.	39%	46%	55%	63%	19%	45%








# Ideas and Combination

- ▶ 数据  
Signed Graph Analysis for the Interpretation of Voting Behavior [6]
- ▶ 算法  
Signed network embedding in social media. [7]





# 参考资料 I

-  伊斯利, et al. 网络, 群体与市场: 揭示高度互联世界的行为原理与效应机制. 清華大學出版社, 2011.
-  Hamilton, Will, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. Advances in Neural Information Processing Systems. 2017.
-  Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014.
-  Traag, Vincent A., and Jeroen Bruggeman. Community detection in networks with positive and negative links. Physical Review E 80.3 (2009): 036115.
-  Kipf, Thomas N., and Max Welling. Semi-supervised classification with graph convolutional networks.



## 参考资料 II

-  Arinik, Nejat, Rosa Figueiredo, and Vincent Labatut. Signed Graph Analysis for the Interpretation of Voting Behavior.
-  Wang, Suhang, et al. Signed network embedding in social media. Proceedings of the 2017 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2017.

