CMSC5741 Big Data Tech. & Apps.

Lecture 10: Mining Social-Network Graphs

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Prof. Michael R. Lyu

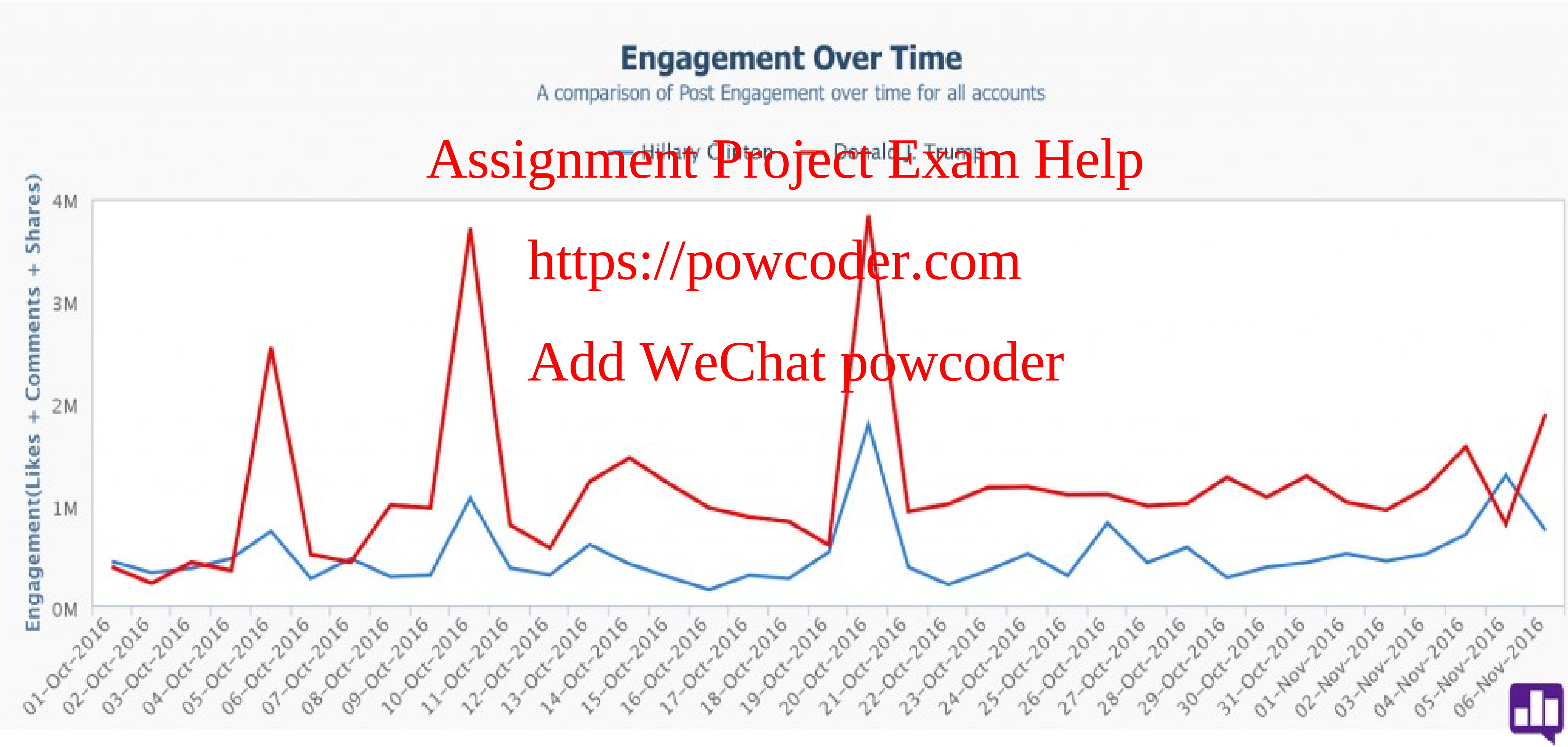
Computer Science & Engineering Dept.

The Chinese University of Hong Kong

Prediction of 2016 US Election

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Social Network Prediction of 2016 US Election



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Outline

- Introduction
- Social Networks as Graph
  - Basics of Graph Theory
  - Important properties of social networks as graphs
  - Analysis of a real-world social network
- Community detection algorithms
  - Girvan-Newman algorithm
  - Spectral clustering
- Conclusions

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Outline

- Introduction
- Social Networks as Graph
  - Basics of Graph Theory
  - Important properties of social networks as graphs
  - Analysis of a real-world social network
- Community detection algorithms
  - Girvan-Newman algorithm
  - Spectral clustering
- Conclusions

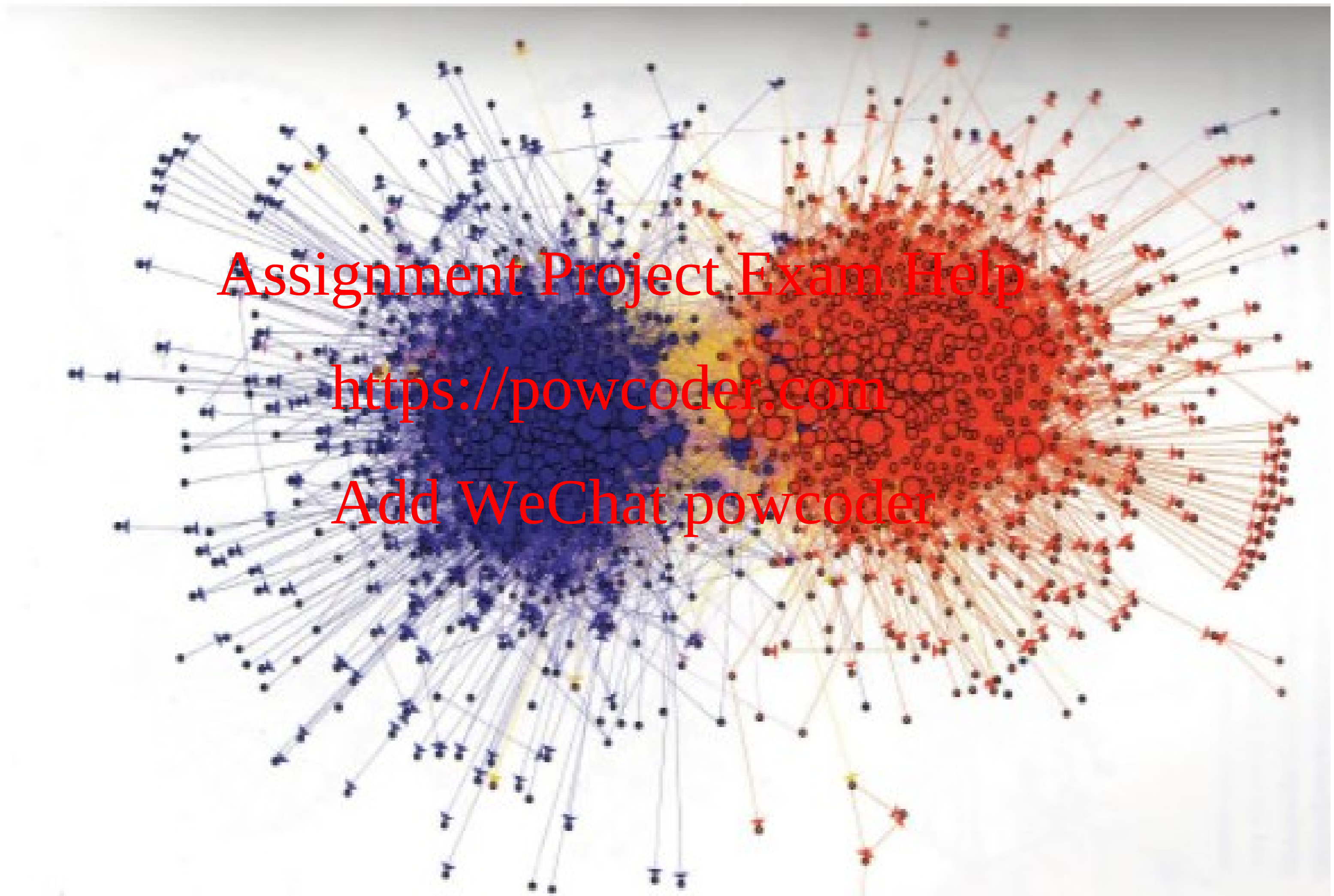
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Networks: Communication

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Graph of the Internet (Autonomous Systems)
Power-law degrees [Faloutsos-Faloutsos-Faloutsos, 1999]
Robustness [Doyle-Willinger, 2005]

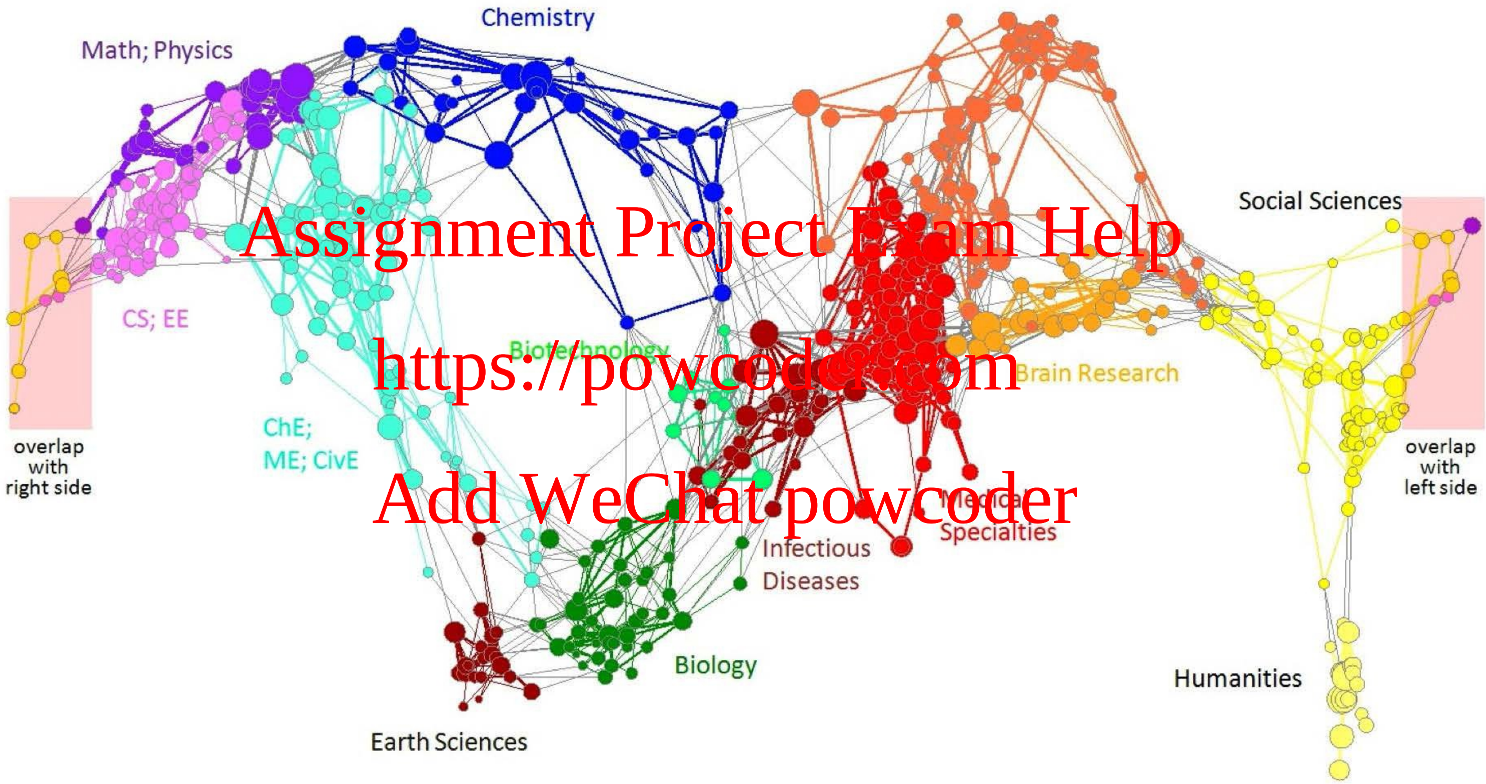
Networks: Media



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Connections between political blogs
Polarization of the network [Adamic-Glance, 2005]

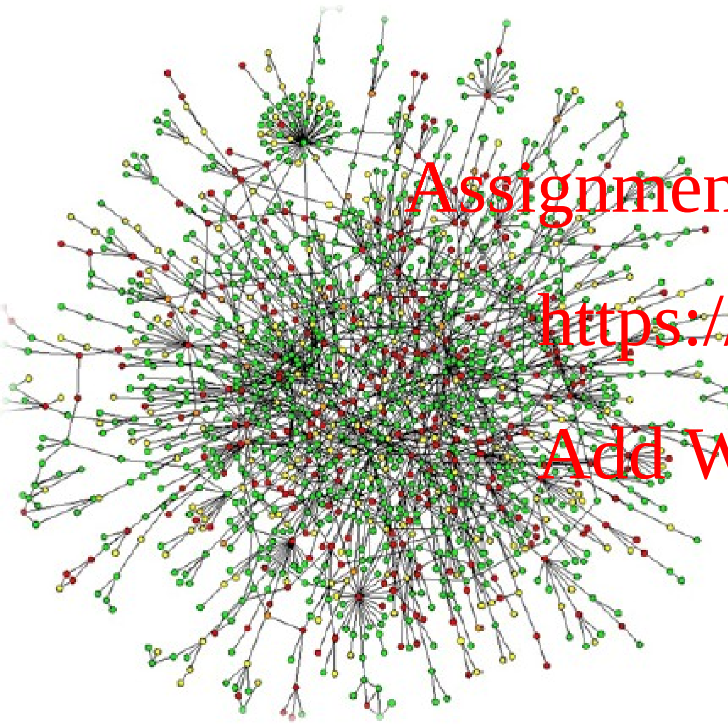
Network: Science

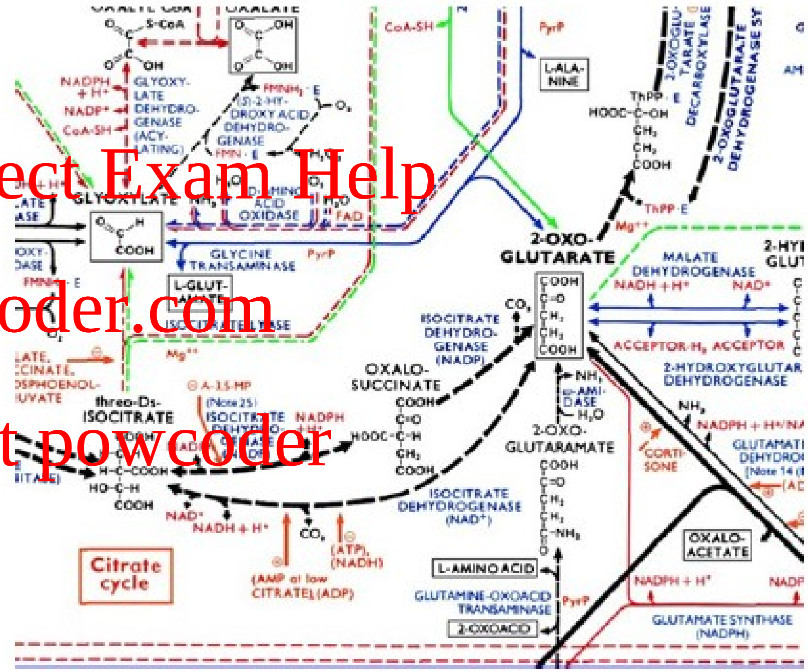


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Citation networks and Maps of science
[Börner et al., 2012]

Networks: Biology



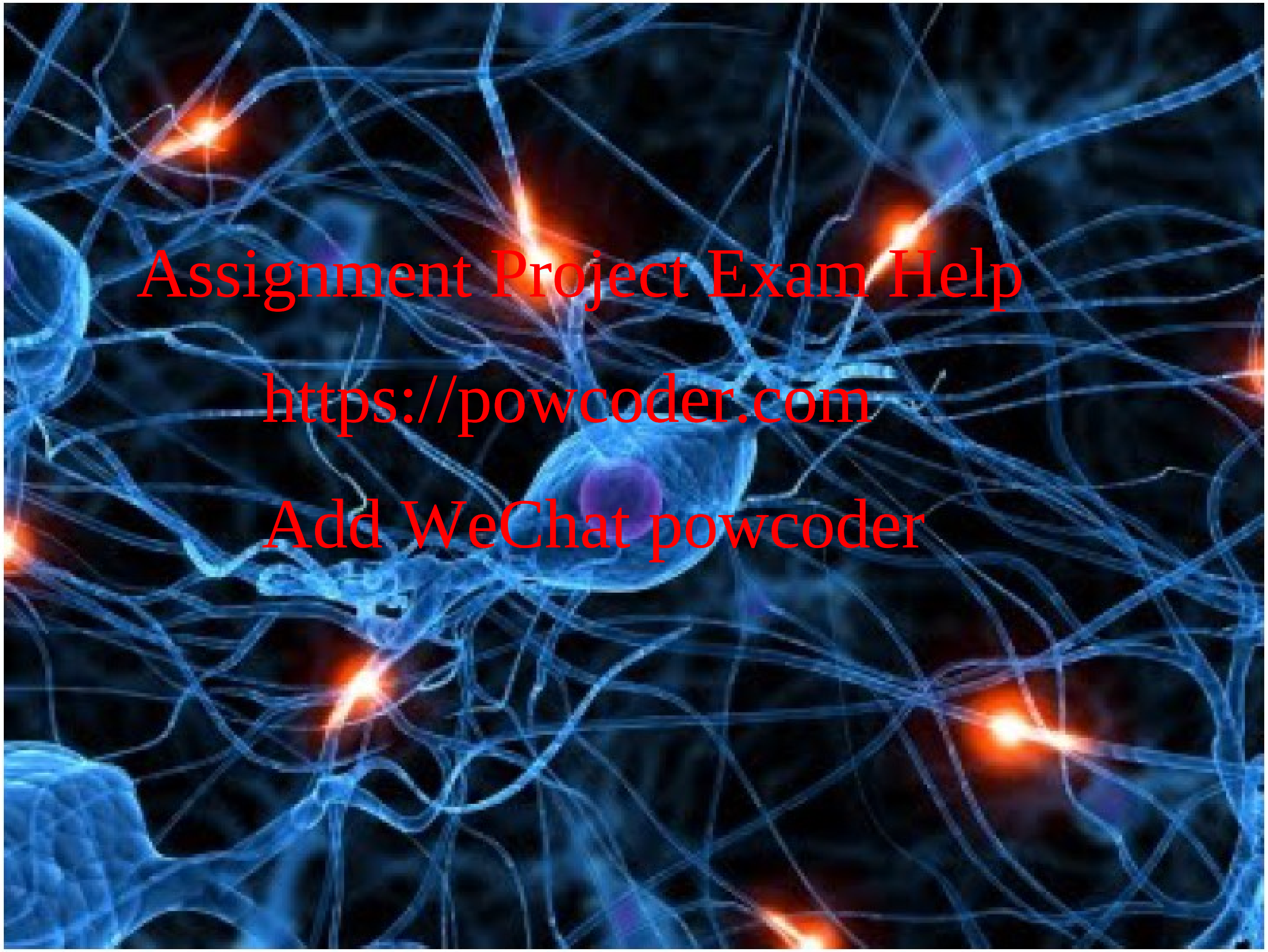


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Protein-Protein Interaction Networks:
Nodes: Proteins
Edges: ‘physical’ interactions

Metabolic networks:
Nodes: Metabolites and enzymes Edges: Chemical reactions

Networks: Brain


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Human brain has between 10-100 billion neurons [Sporns, 2011]

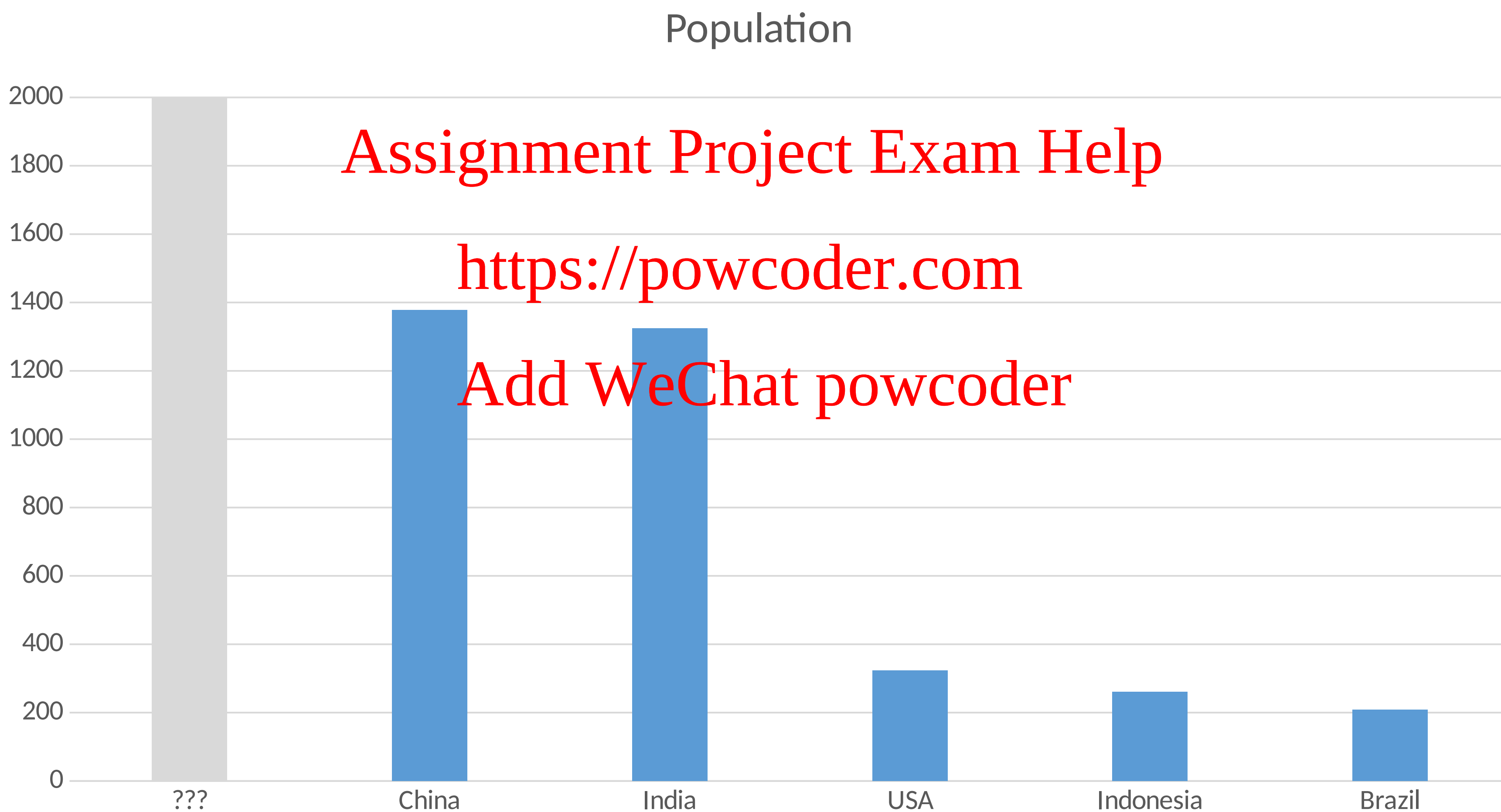
Networks: Social

Assignment Project Exam Help

https://powcoder.com

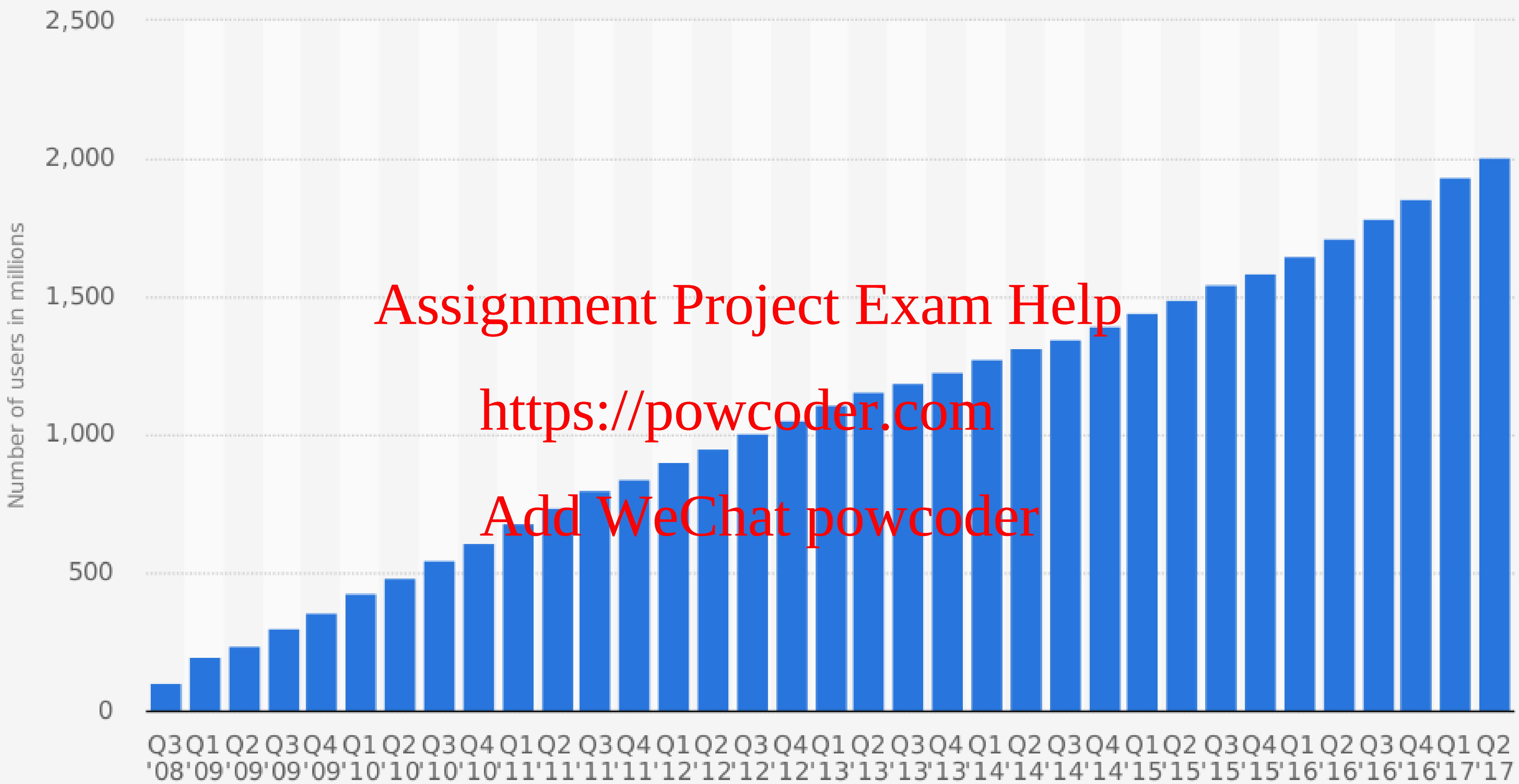
Add WeChat powcoder

Country with Largest World Population?



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Number of monthly active Facebook users worldwide as of 2nd quarter 2017 (in millions)

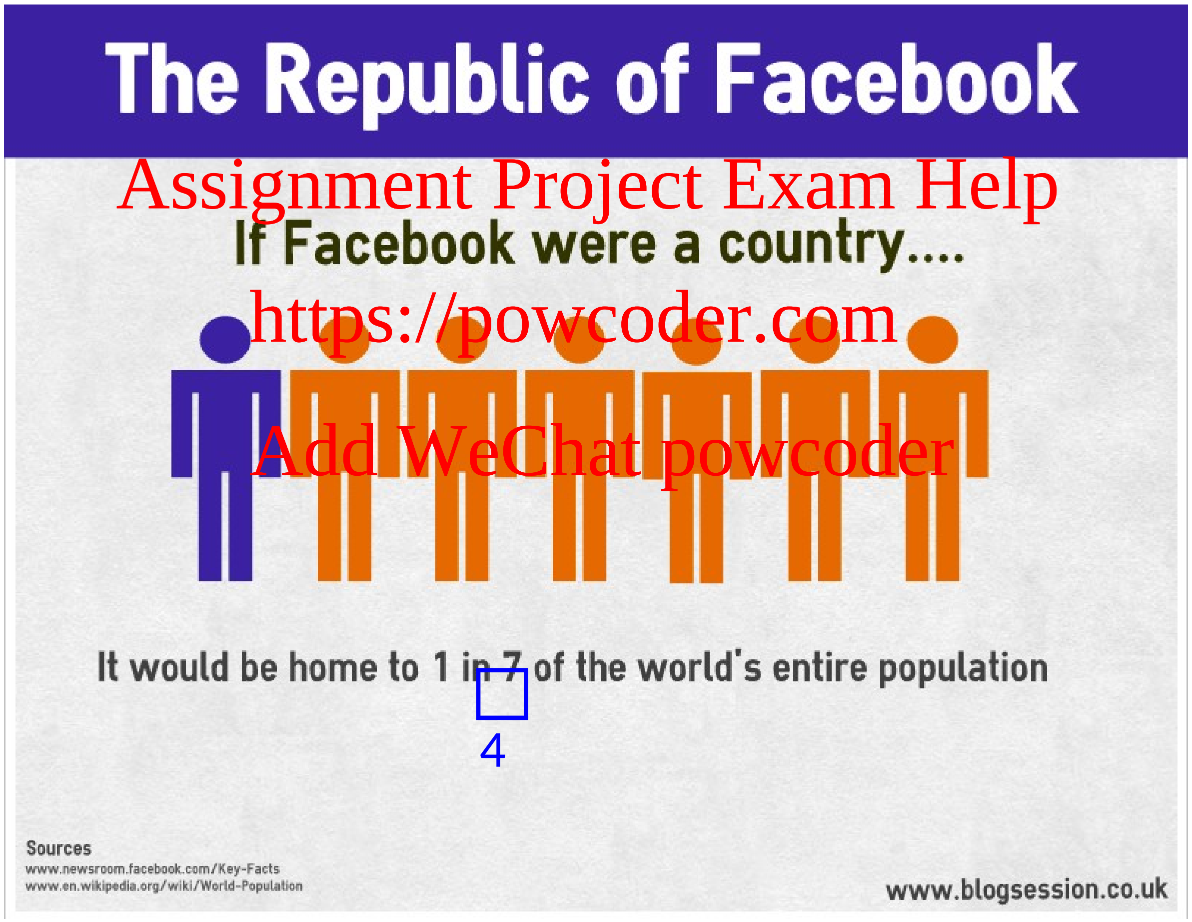


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Source
Facebook
© Statista 2017

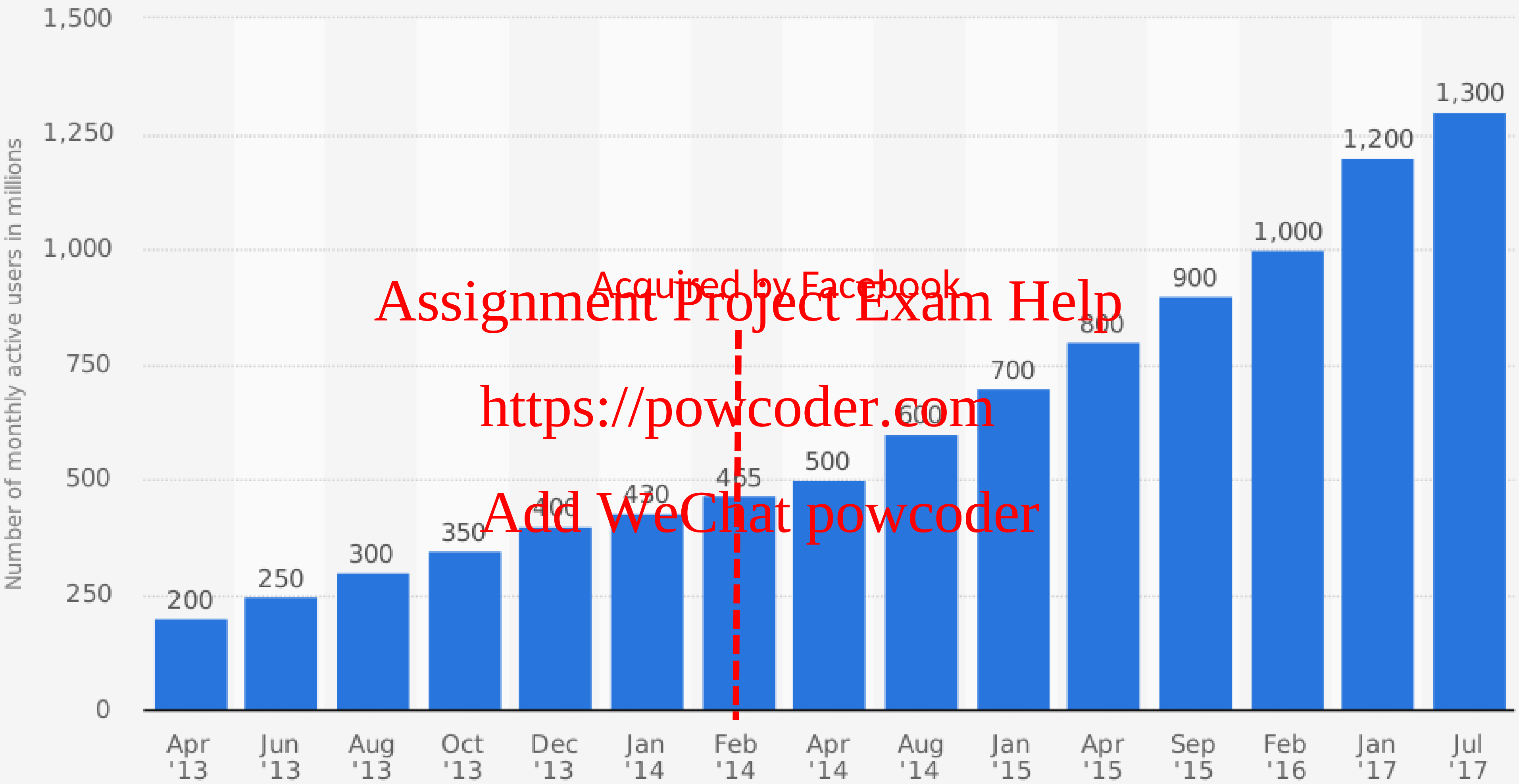
Additional Information:
Worldwide; Facebook; 3rd quarter 2008 to 2nd quarter 2017

If Facebook were a Country..



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Number of monthly active WhatsApp users worldwide from April 2013 to July 2017 (in millions)



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Source
WhatsApp; Facebook; TechCrunch
© Statista 2017

Additional Information:
Worldwide; WhatsApp; April 2013 to July 2017

statista

Networks!

- Behind each such system there is an intricate wiring diagram, **a network**, that defines the **interactions** between the components

Assignment Project Exam Help

https://powcoder.com

- **We will never understand these systems unless we understand the networks behind them!**

Add WeChat powcoder

Social Web - The Lab of Knowledge ar

Assignment Project Exam Help

The web is our "laboratory" for knowledge and understanding humanity

https://powcoder.com

Add WeChat powcoder

Networks: Impact

licensed under (cc) Attribution-NonCommercial-ShareAlike 2.0 Germany | Ludwig Gatzke | http://flickr.com/photos/stabilo-boss/

Cisco Market cap: $211b billion (2y ago it was 150b)

Google Market cap: $740b billion (2y ago it was 536b)

Facebook Market cap: $400b billion (2y ago it was 336b)

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Networks: Problems

- **Social networks:**
  - Link prediction, friend recommendation
  - Social circle detection, community detection
  - Social recommendations
  - Identifying influential nodes, viral marketing
- **Communication networks:**
  - Intrusion detection, fraud
  - Churn prediction
- **Information networks:**
  - Navigational aids

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

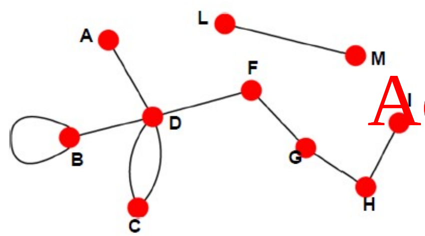
Outline

- Introduction
- Social Networks as Graph
  - Basics of Graph Theory
  - Important properties of social networks as graphs
  - Analysis of a real-world social network
- Community detection algorithms
  - Girvan-Newman algorithm
  - Spectral clustering
- Conclusions

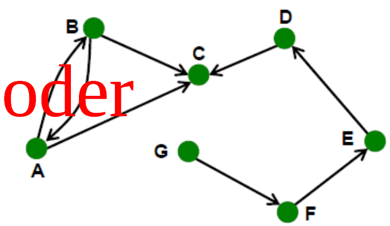
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Social Network as Graph

- **Undirected graphs**
  - **Links:** undirected (symmetrical, reciprocal relations)



- **Directed graphs**
  - **Links:** directed (asymmetrical relations)

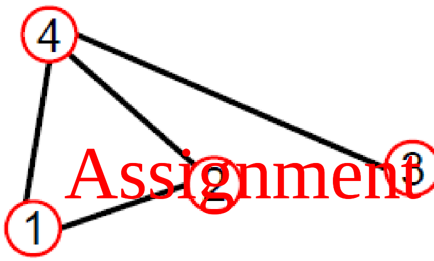


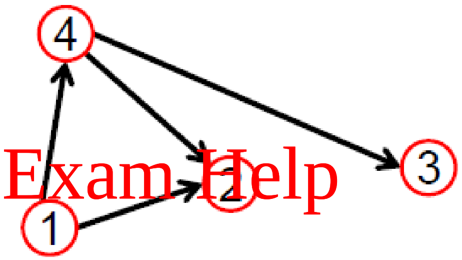
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- Undirected links:
  - Collaborations
  - Friendship on Facebook

- Directed links:
  - Phone calls
  - Following on Twitter

Adjacency Matrix





Assignment Project Exam Help

https://powcoder.com

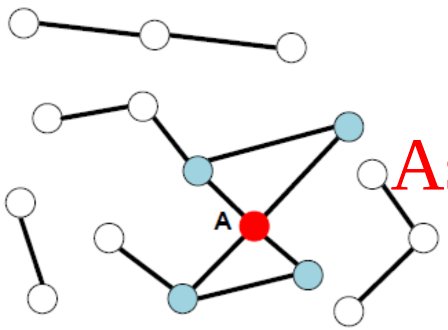
$A_{ij} = 1$ if there is a link from node ***i*** to node ***j***

$A_{ij} = 0$ otherwise

Add WeChat powcoder

Node Degrees

Undirected



- Node degree : the number of edges adjacent to node

Assignment Project Exam Help

https://powcoder.com

  - Avg. degree: $\bar{k} = \langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2E}{N}$

Add WeChat powcoder

- In directed networks we define an **in-degree** and **out-degree**.
  - The (total) degree of a node is the sum of in- and out-degrees.

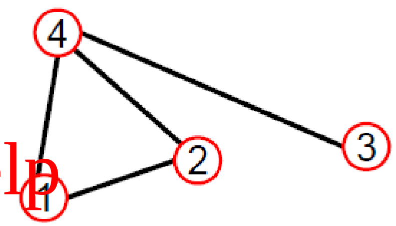
$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

$$\bar{k} = \frac{E}{N} \qquad \overline{k^{in}} = \overline{k^{out}}$$

Degree Matrix

- This graph has entries only on the diagonal.
- The entry for row and column *i* is the degree of the *i*th node

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



$$D = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Laplacian Matrix

- Suppose a graph as adjacency matrix and degree matrix .

Assignment Project Exam Help

- Laplacian matrix is defined as .

https://powcoder.com

- the Laplacian matrix
  - Has the same entries as on the diagonal.

Add WeChat powcoder

  - Off the diagonal, at row and column ,
    - has −1 if there is an edge between nodes and
    - 0 if not.

In-class Practice 1

- Go to practice

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Complete Graph

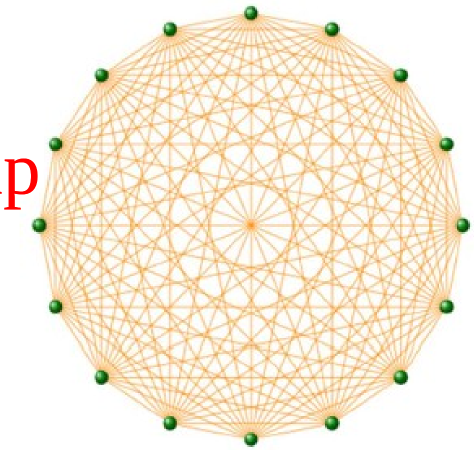
- The **maximum number of edges** in an undirected graph on ***N*** nodes is

Assignment Project Exam Help

$$E_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$$

https://powcoder.com

Add WeChat powcoder



- A graph with the number of edges is a **complete graph**
  - and its average degree is
  - ***N-1***

Networks are Sparse Graphs

- **Most real-world networks are sparse**
  - **$E \ll E_{max}$ (or $k \ll N-1$)**

Assignment Project Exam Help

WWW (Stanford-Berkeley):	N=319,717	⟨k⟩=9.65
Social networks (LinkedIn):	N=6,946,668	⟨k⟩=8.87
Communication (MSN IM):	N=242,720,596	⟨k⟩=11.1
Coauthorships (DBLP):	N=317,080	⟨k⟩=6.62
Internet (AS-Skitter):	N=1,719,037	⟨k⟩=14.91
Roads (California):	N=1,957,027	⟨k⟩=2.82
Proteins (S. Cerevisiae):	N=1,870	⟨k⟩=2.39

https://powcoder.com
Add WeChat powcoder

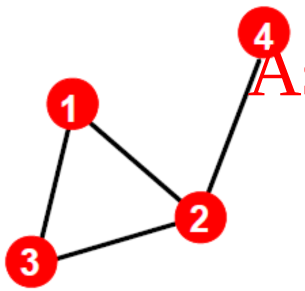
(Source: *Leskovec et al., Internet Mathematics, 2009*)

- **Consequence: Adjacency matrix is filled with zeros!**
  - **Density of the matrix (E/N^2):**

WWW=$1.51 * 10^{-5}$, MSN IM Instant Messenger = $2.27 * 10^{-8}$

More Types of Graphs

Unweighted



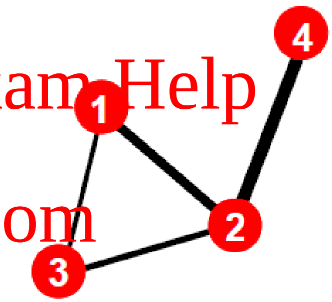
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$E = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad \bar{k} = \frac{2E}{N}$$

Examples: Friendship, Hyperlink

Weighted



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

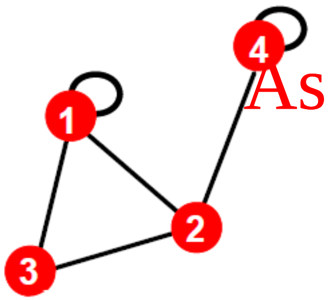
$$E = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad \bar{k} = \frac{2E}{N}$$

Examples: Collaboration, Internet, Roads

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

More Types of Graphs

Self-edges (Self-loops)



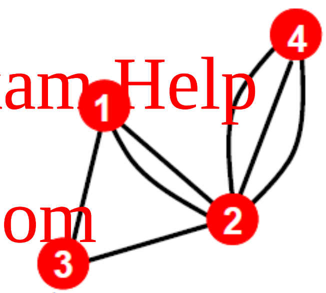
$$A_{ij} = \begin{pmatrix} \boxed{1} & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & \boxed{1} \end{pmatrix}$$

$$A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$$

$$E = \frac{1}{2}\sum_{i,j=1,i\neq j}^{N} A_{ij} + \sum_{i=1}^{N} A_{ii} \qquad ?$$

Examples: Proteins, Hyperlink

Multigraph



$$A_{ij} = \begin{pmatrix} 0 & \boxed{2} & 1 & 0 \\ \boxed{2} & 0 & 1 & \boxed{3} \\ 1 & 1 & 0 & 0 \\ 0 & \boxed{3} & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$E = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad \bar{k} = \frac{2E}{N}$$

Examples: Communication, Collaboration

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Network Representations

WWW >> directed multigraph with self-edges

Assignment Project Exam Help

Facebook friendships >> undirected, unweighted

https://powcoder.com

Citation networks >> unweighted, directed, acyclic

Add WeChat powcoder

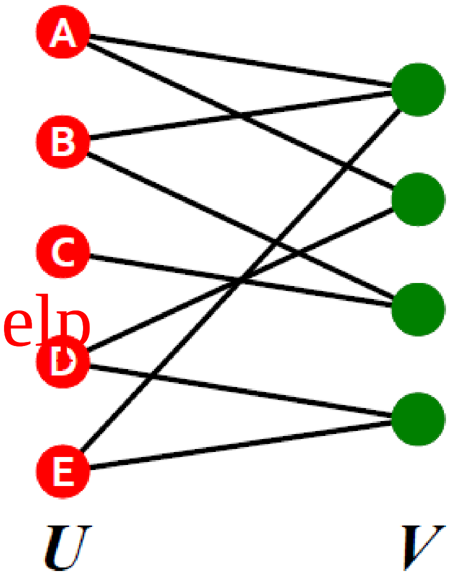
Collaboration networks >> undirected multigraph or weighted graph

Mobile phone calls >> directed, (weighted?) multigraph

Protein Interactions >> undirected, unweighted with self-interactions

Bipartite Graph

- **Bipartite graph** is a graph whose nodes can be divided into two disjoint sets ***U*** and ***V*** such that every link connects a node in ***U*** to one in ***V***; that is, ***U*** and ***V*** are **independent sets**
- **Examples:**
  - Authors-to-papers (they authored)
  - Actors-to-Movies (they appeared in)
  - Users-to-Movies (they rated)
- **"Folded" networks:**
  - Author collaboration networks
  - Movie co-rating networks



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Outline

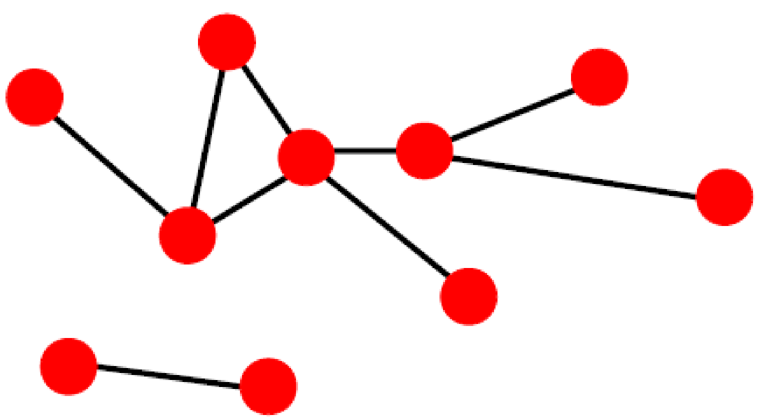
- Introduction
- Social Networks as Graph
  - Basics of Graph Theory
  - Important properties of social networks as graphs
  - Analysis of a real-world social network
- Community detection algorithms
  - Girvan-Newman algorithm
  - Spectral clustering
- Conclusions

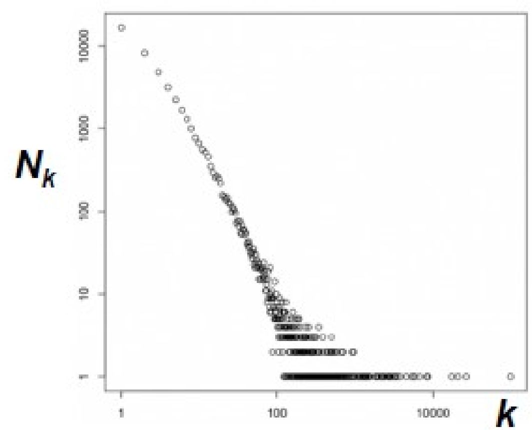
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Degree Distribution

- Degree distribution : Probability that a randomly chosen node has degree ***k***
  - $\boldsymbol{N_k}$ = # nodes with degree ***k***
- Normalized:
  - $P(k) = N_k / N$ ➔ **plot**

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder





Paths in a Graph

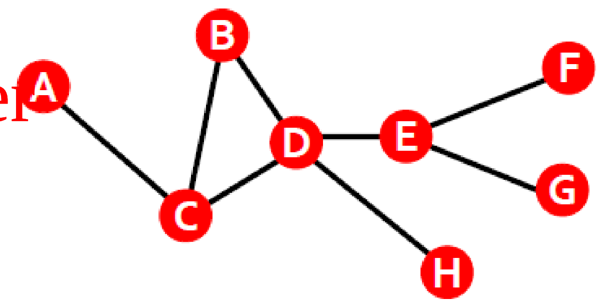
- A ***path*** is a sequence of nodes in which each node is linked to the next one

Assignment Project Exam Help

$P_n = \{i_0, i_1, i_2, \ldots, i_n\}$ $P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$

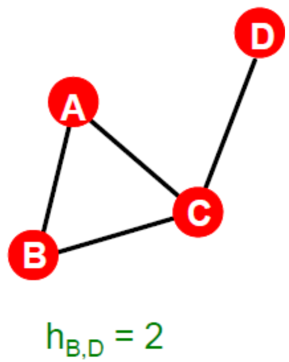
https://powcoder.com

Add WeChat powcoder



- Path can intersect itself and pass through the same edge multiple times
  - E.g.: ACBDCDEG
  - In a directed graph a path can only follow the direction of the "arrow"

Distance in a Graph



- **Distance (shortest path, geodesic)** between a pair of nodes is defined as the number of edges along the shortest path connecting the nodes
  - *If the two nodes are disconnected, the distance is usually defined as infinite
- In **directed graphs** paths need to follow the direction of the arrows
  - Consequence: Distance is **not symmetric**: $h_{A,C} \neq h_{C,A}$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Network Diameter

- **Diameter**: the maximum (shortest path) distance between any pair of nodes in a graph

Assignment Project Exam Help

- **Average path length** for a connected graph (component) or a strongly connected (component of a) directed graph:

https://powcoder.com

Add WeChat powcoder

$$\bar{h} = \frac{1}{2E_{\max}} \sum_{i,j \neq i} h_{ij}$$

$$E_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$$

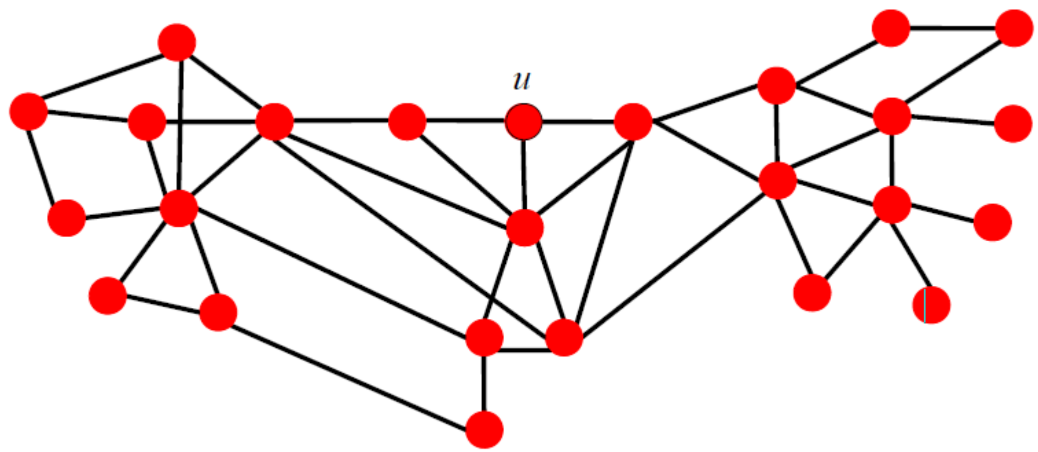
where h_{ij} is the distance from node i to node j

  - Many times we compute the average only over the connected pairs of nodes (we ignore "infinite" length paths)

Finding the Shortest Path

- **Breath‑First Search:**
  - Start with node u, mark it to be at distance $h_u(u)=0$, add u to the queue
  - While the queue not empty:
    - Take node v off the queue, put its unmarked neighbors w into the queue and mark $h_u(w)=h_u(v)+1$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Clustering Coefficient

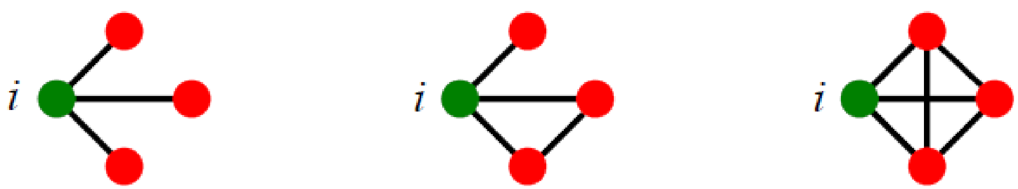
- **Clustering coefficient:**
  - What portion of ***i***'s neighbors are connected?
  - Node ***i*** with degree k_i

Assignment Project Exam Help

https://powcoder.com

where e_i is the number of edges between the neighbors of node i

Add WeChat powcoder



Clustering Coefficient

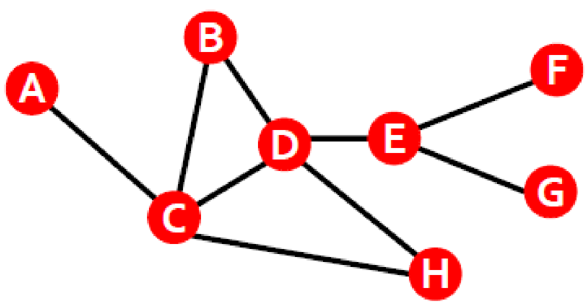
- **Clustering coefficient:**
  - What portion of ***i***'s neighbors are connected?
  - Node ***i*** with degree $\boldsymbol{k_i}$

Assignment Project Exam Help

https://powcoder.com

where e_i is the number of edges between the neighbors of node i

Add WeChat powcoder



$k_B=2,\ \ e_B=1,$

$k_D=4,\ \ e_D=2,$

Key Network Properties

- Degree distribution:
- Path length:
- Clustering coefficient:

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Outline

- Introduction
- Social Networks as Graph
  - Basics of Graph Theory
  - Important properties of social networks as graphs
  - Analysis of a real-world social network
- Community detection algorithms
  - Girvan-Newman algorithm
  - Spectral clustering
- Conclusions

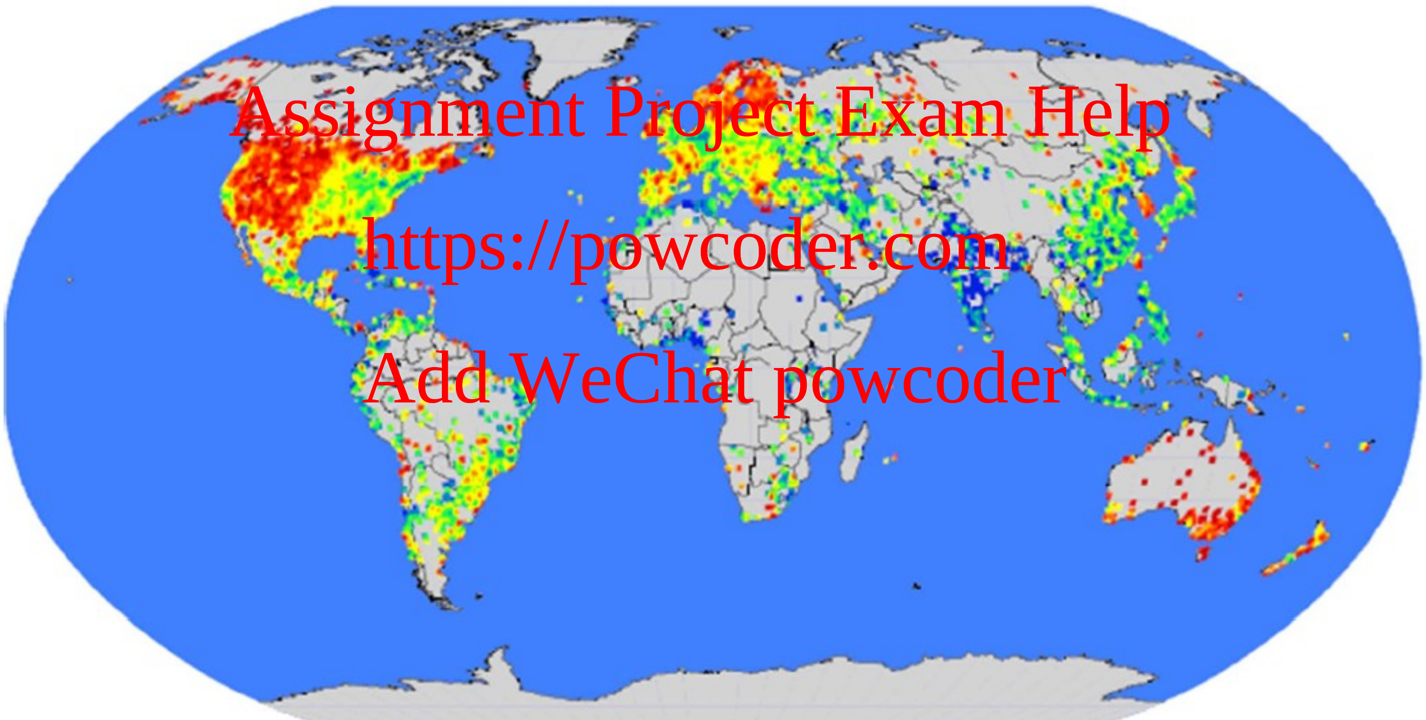
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

The MSN Messenger

- **MSN Messenger activity in June 2006:**
  - 150Gb/day (compressed)
  - 4.5Tb / month
  - 245 million users logged in
  - 180 million users engaged in conversations
  - More than 30 billion conversations
  - More than 255 billion exchanged messages

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Communication: Geography



Assignment Project Exam Help

https://powcoder.com

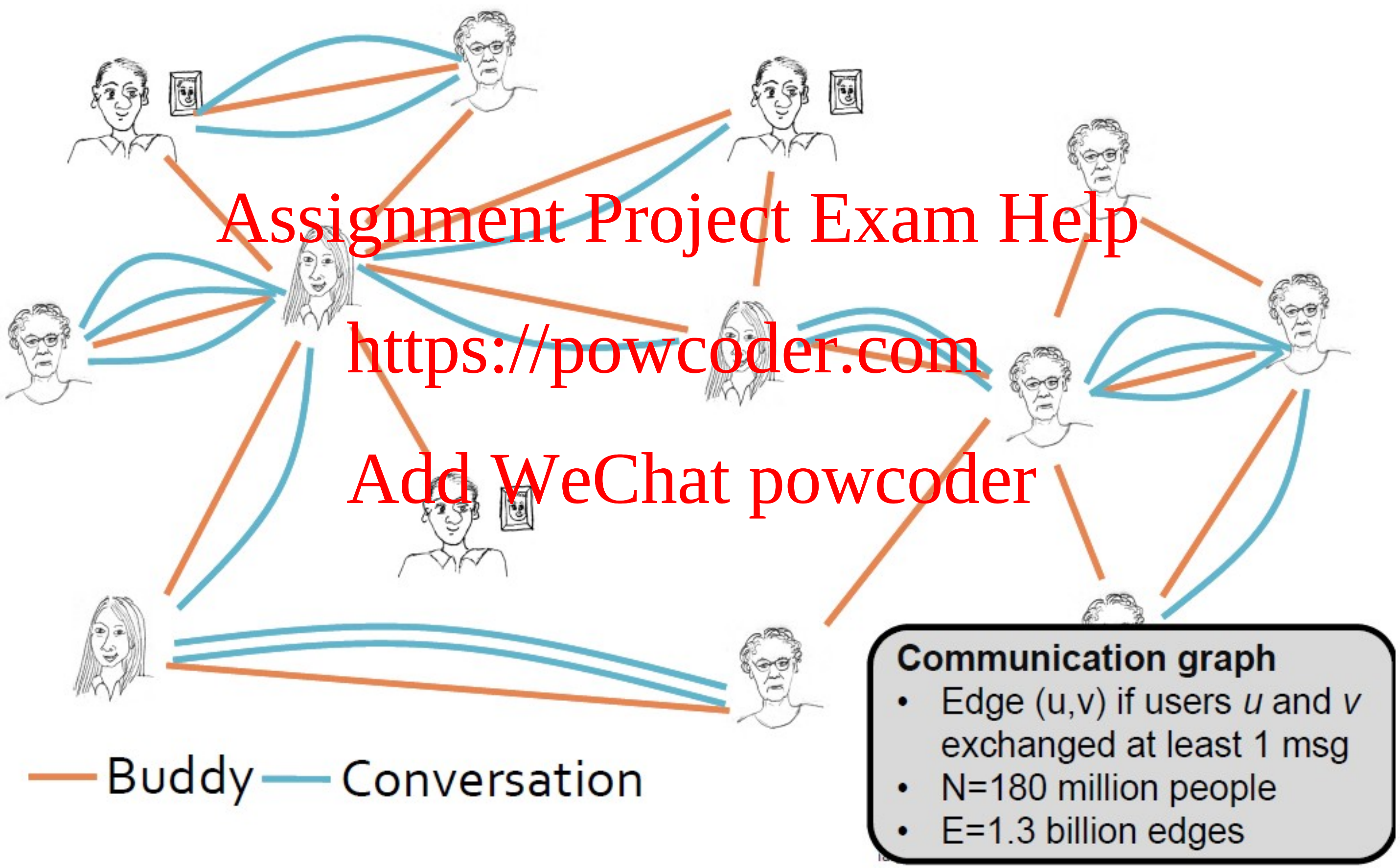
Add WeChat powcoder

Communication Network

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

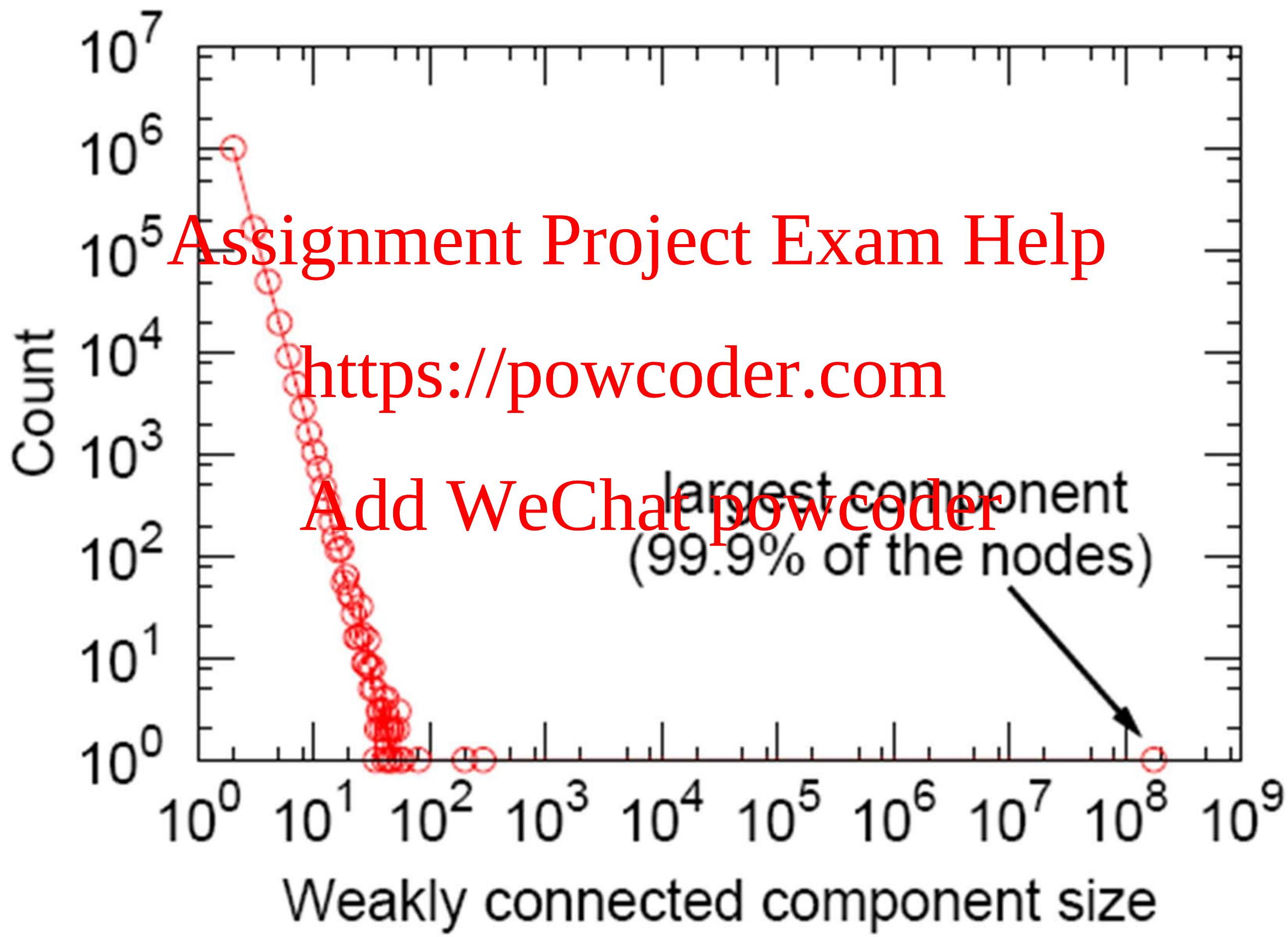
Network: 180M people, 1.3B edges

Messaging as a Network



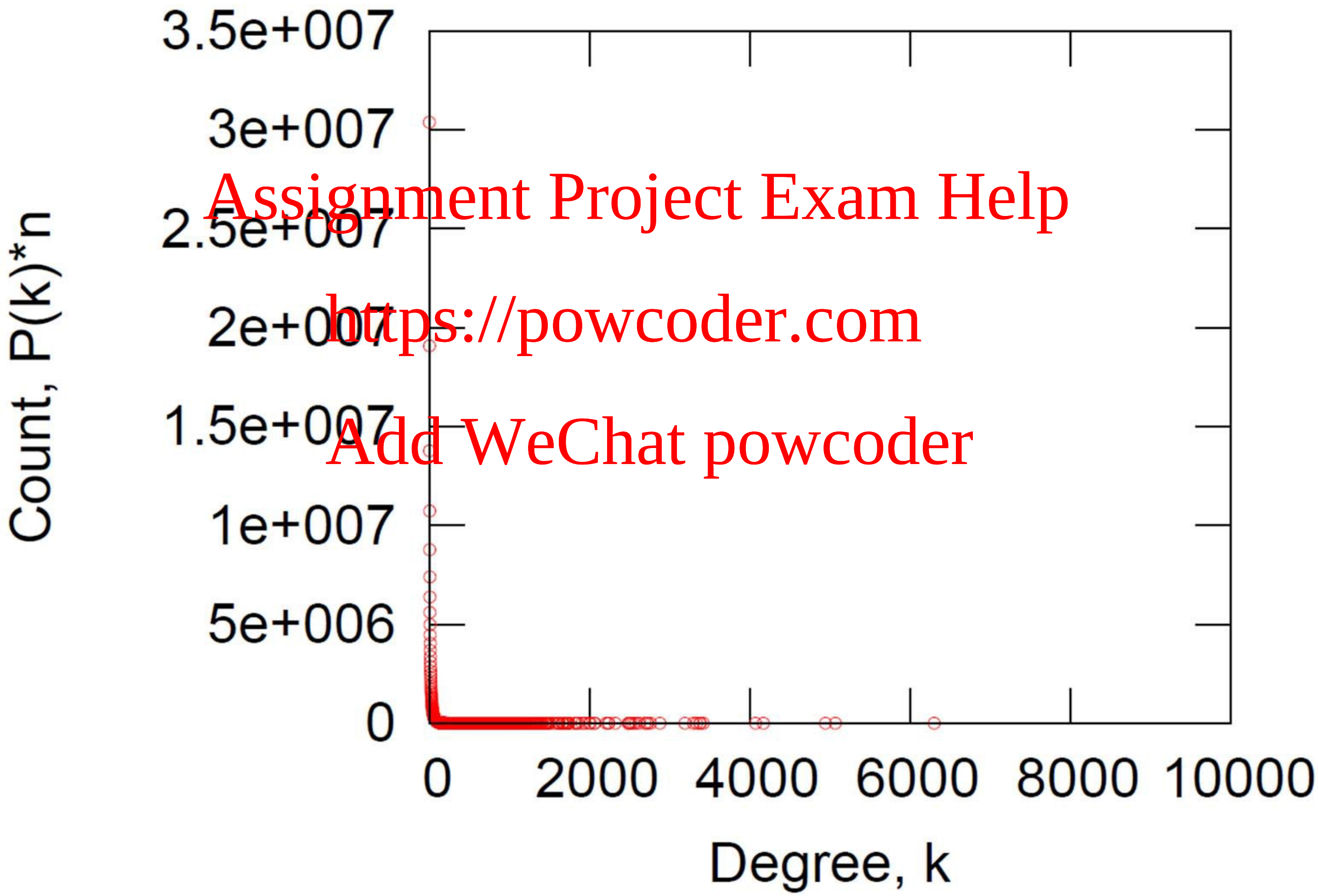
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

MSN Network: Connectivity



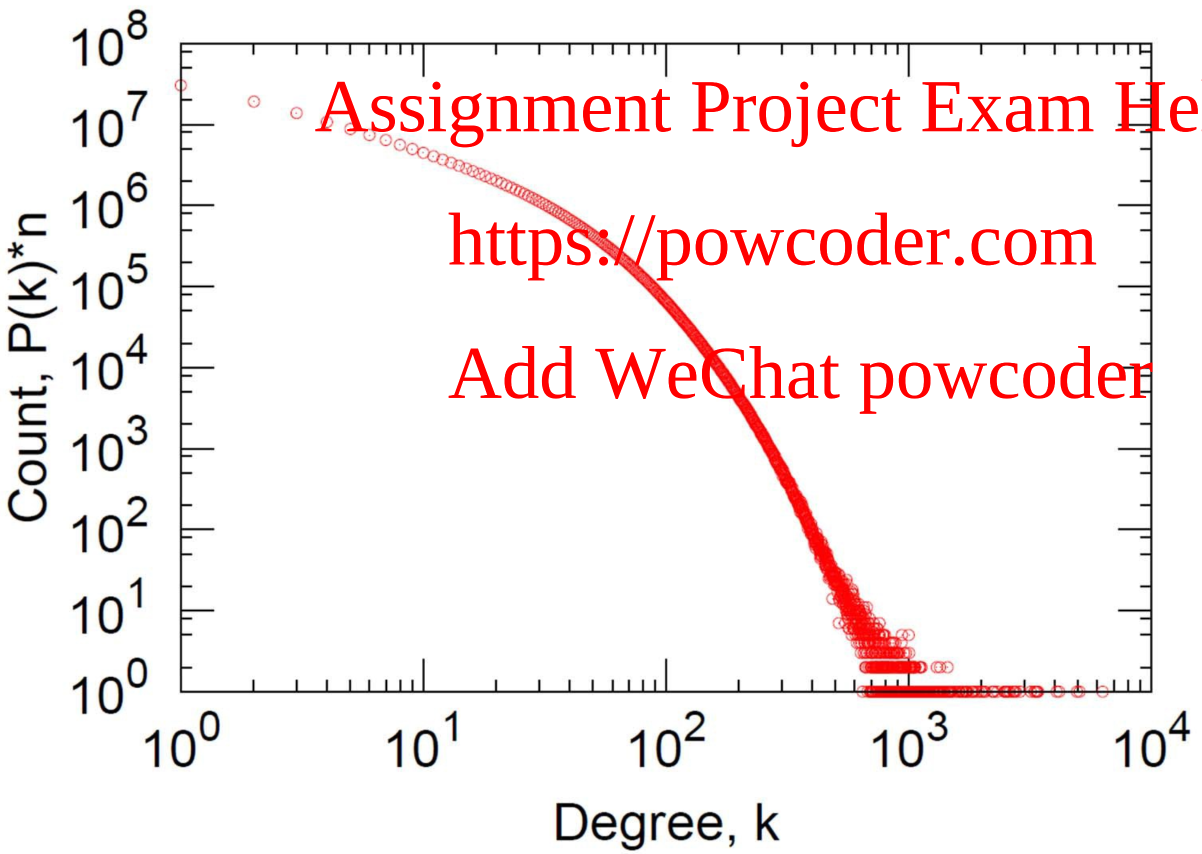
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

MSN: Degree Distribution



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

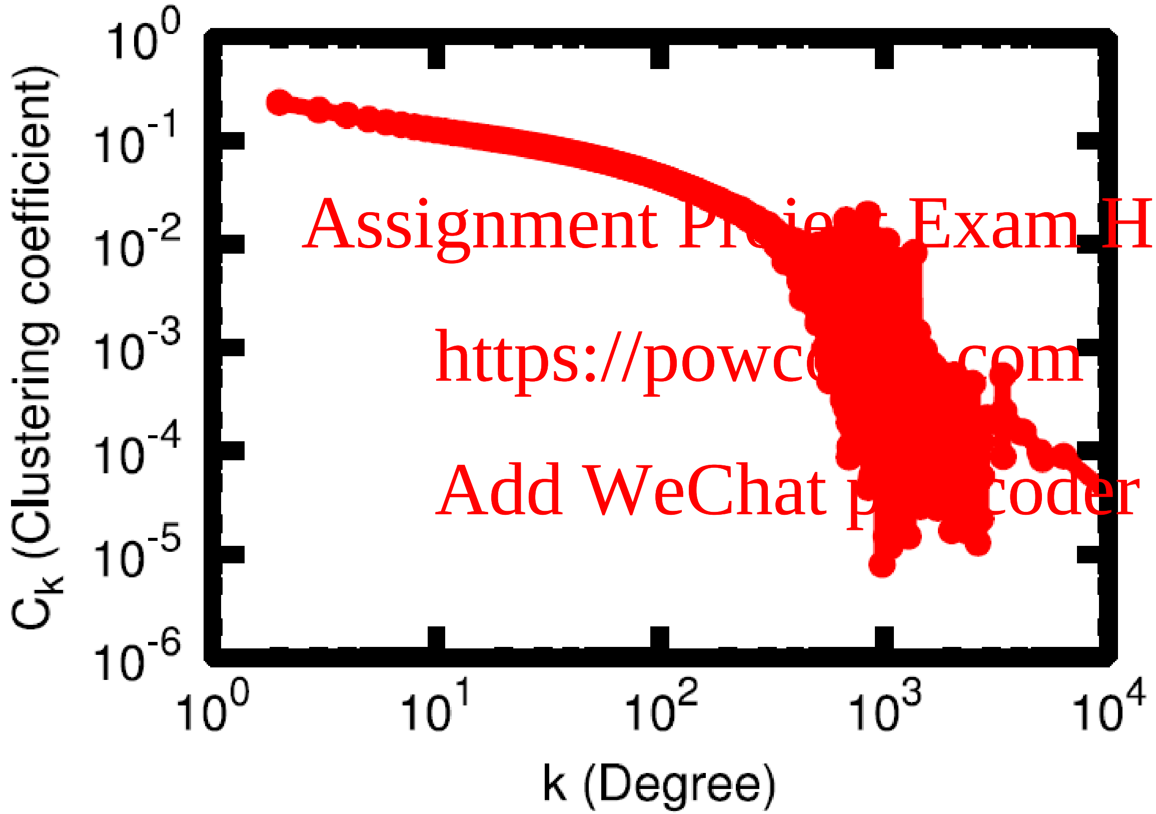
MSN: Log-Log Degree Distribution



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

We plot the same data as on the previous slide, just the axes are now logarithmic.

MSN: Clustering

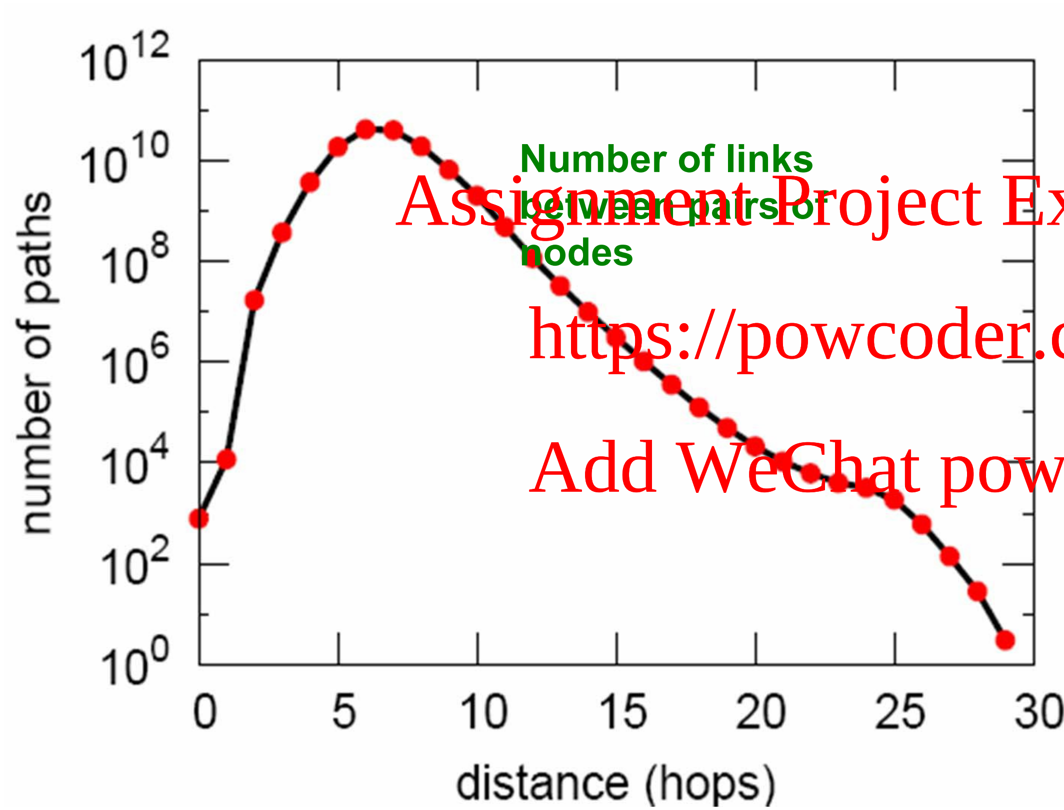


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Avg. clustering of the MSN:
C = 0.1140

C_k: average C_i of nodes i of degree k: $C_k = \frac{1}{N_k} \sum_{i:k_i=k} C_i$

MSN: Diameter



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Avg. path length **6.6**
90% of the people can be reached in < 8 hops

nodes as we do BFS out of a random node

Steps	#Nodes
0	1
1	10
2	78
3	3,96
4	8,648
5	3,299,252
6	28,395,849
7	79,059,497
8	52,995,778
9	10,321,008
10	1,955,007
11	518,410
12	149,945
13	44,616
14	13,740
15	4,476
16	1,542
17	536
18	167
19	71
20	29
21	16
22	10
23	3
24	2
25	3

MSN: Key Network Properties

- Degree distribution: heavily skewed. avg degree = 14.4
- Path length: 6.6
- Clustering coefficient: 0.11

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

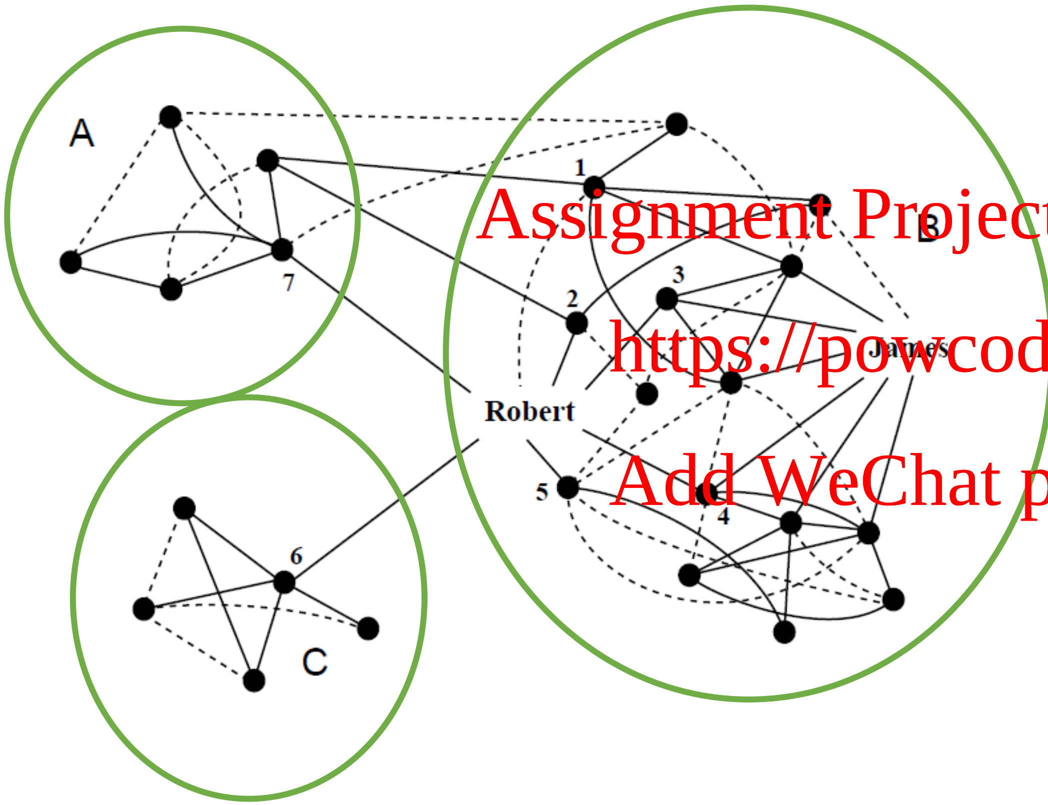
Outline

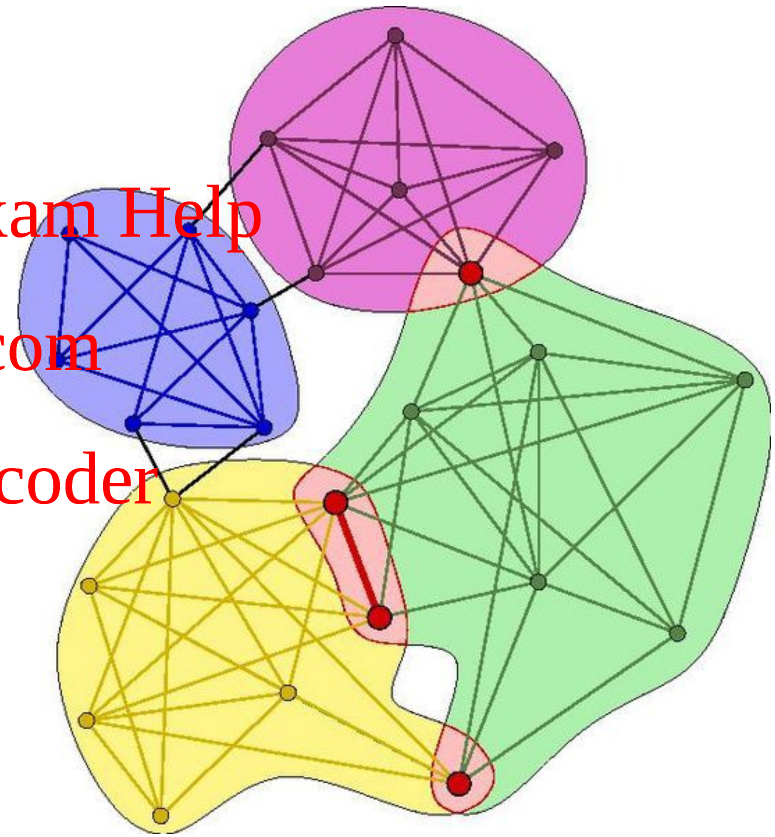
- Introduction
- Social Networks as Graph
  - Basics of Graph Theory
  - Important properties of social networks as graphs
  - Analysis of a real-world social network
- Community detection algorithms
  - Girvan-Newman algorithm
  - Spectral clustering
- Conclusions

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Community Detection

How to find communities?





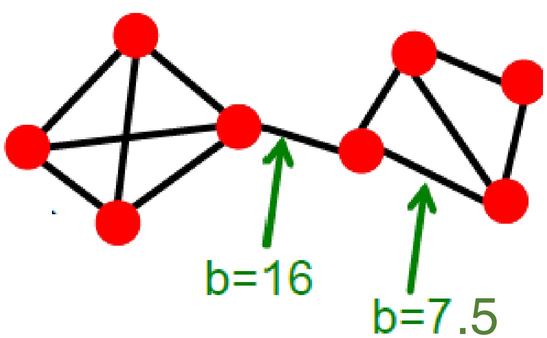
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

We will work with **undirected** (unweighted) networks

Betweenness: Strength of Ties

- **Edge betweenness**: number of shortest paths (among all pair of vertices) passing over the edge.
- If there is more than one shortest path between a pair of nodes, each path is assigned equal weight such that the total weight of all the paths is equal to 1.

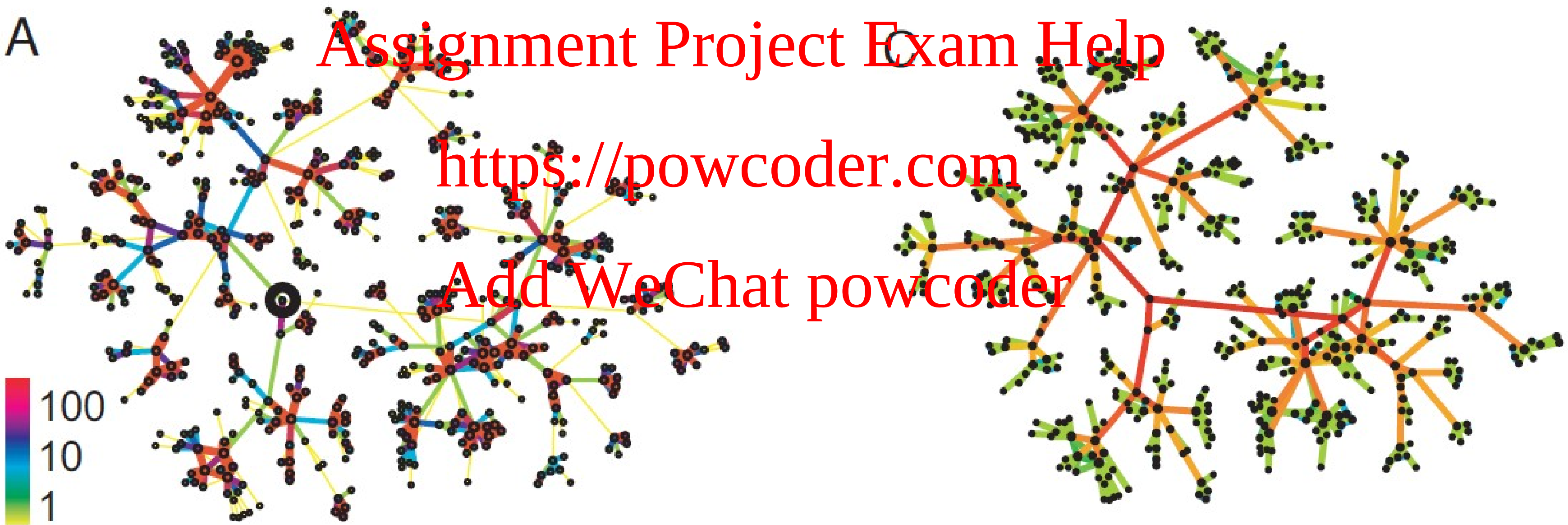
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Simple algorithm for calculating betweeness.

1. Repeat for each vertex
   a) compute its shortest path to all other vertices.
   b) for each edge, count the number of the shortest paths passing over the edge
2. Sum up the results from each vertex.

Betweenness: Strength of Ties



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Calculation for 1 vertex

Calculation for all vertices

In-class Practice 2

- Go to practice

Assignment Project Exam Help

https://powcoder.com

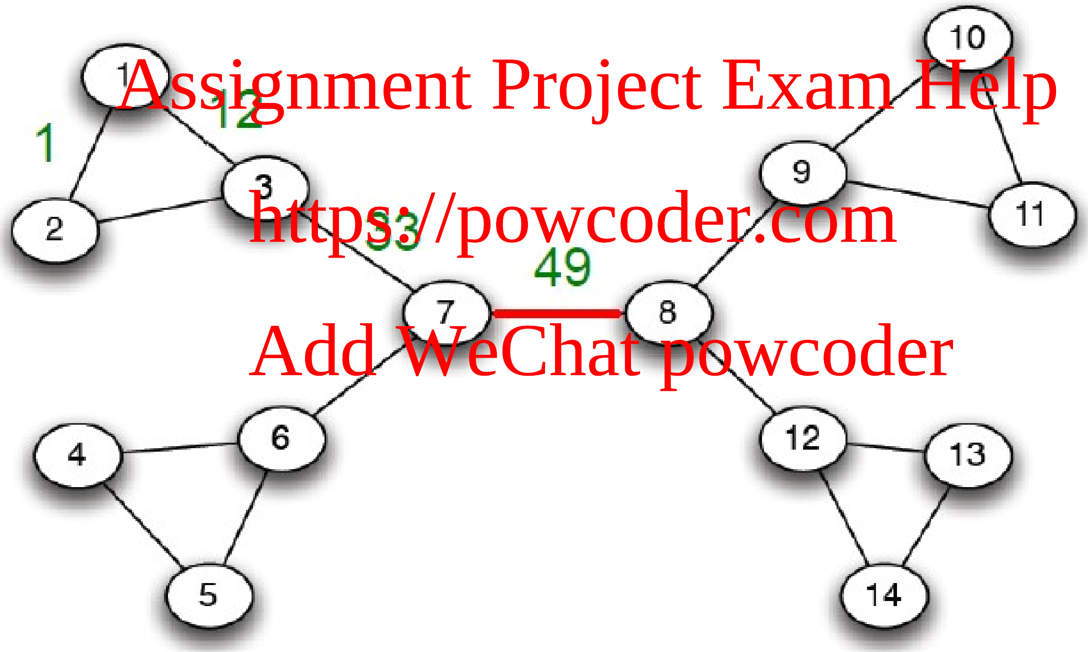
Add WeChat powcoder

Method for Community Detection: Girvan-Newman

- Divisive hierarchical clustering based on the notion of edge **betweenness**:
  - **Number of shortest paths passing through the edge**
- **Girvan-Newman Algorithm:**
  - **Undirected unweighted networks**
  - **Repeat until no edges are left:**
    - Calculate betweenness of edges
    - Remove edges with highest betweenness
  - Connected components are communities
  - Gives a hierarchical decomposition of the network

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Girvan-Newman: Example

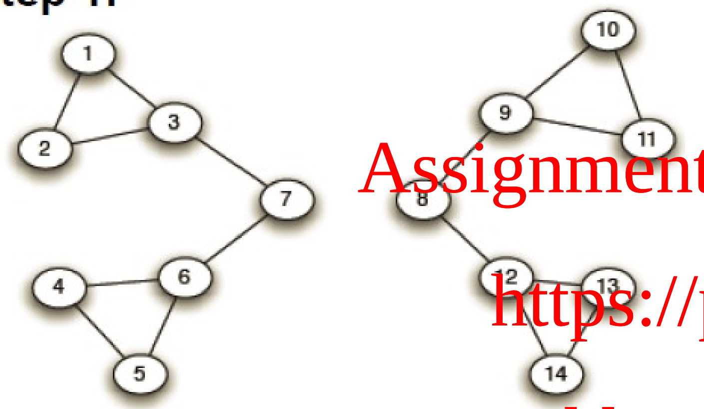


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

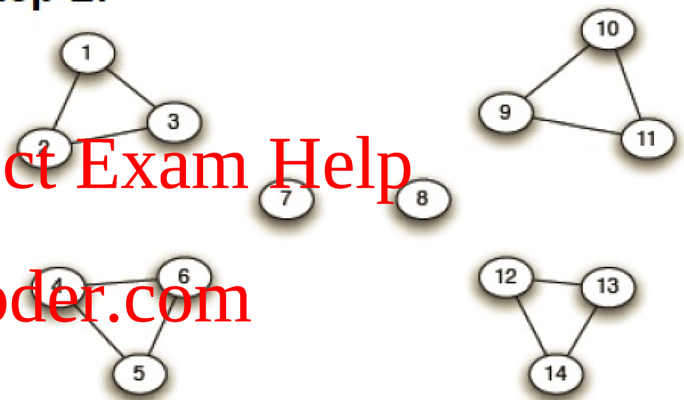
Need to re-compute betweenness at every step

Girvan-Newman: Example

Step 1:



Step 2:

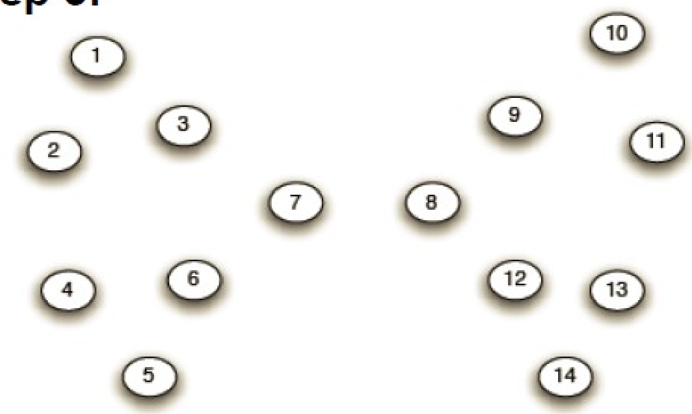


Assignment Project Exam Help

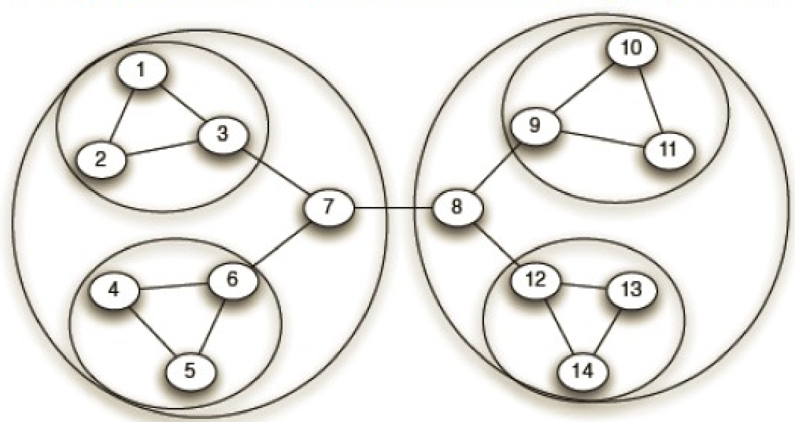
https://powcoder.com

Add WeChat powcoder

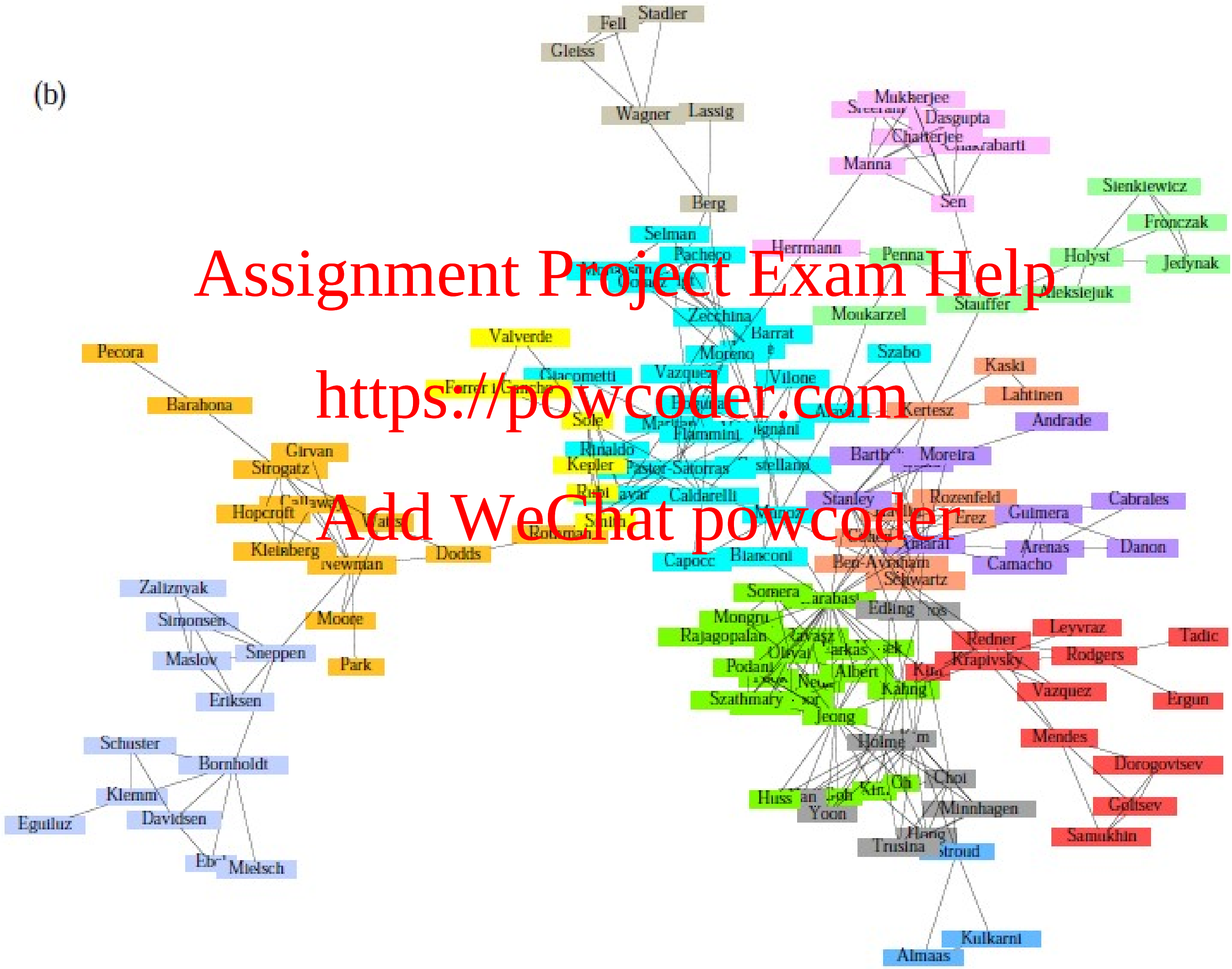
Step 3:



Hierarchical network decomposition:



Girvan-Newman



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Communities in physics collaborations

Problems of Girvan-Newman

- Computing betweenness is slow.
- No theoretical guarantees.
- Not widely applied.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Outline

- Introduction
- Social Networks as Graph
  - Basics of Graph Theory
  - Important properties of social networks as graphs
  - Analysis of a real-world social network
- Community detection algorithms
  - *Girvan-Newman* algorithm
  - Spectral clustering
- Conclusions

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Graph Partitioning

- **Undirected graph $G\ (V,E)$:**

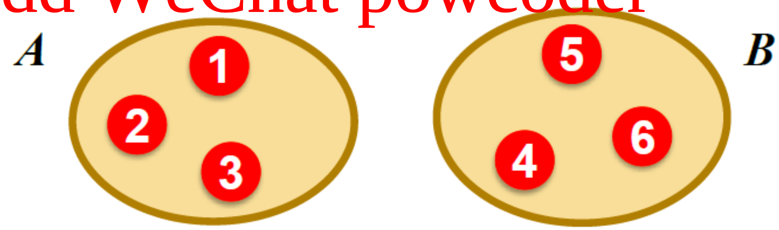
Assignment Project Exam Help

- **Bi-partitioning task:**

https://powcoder.com

  - Divide vertices into two disjoint groups A, B

Add WeChat powcoder

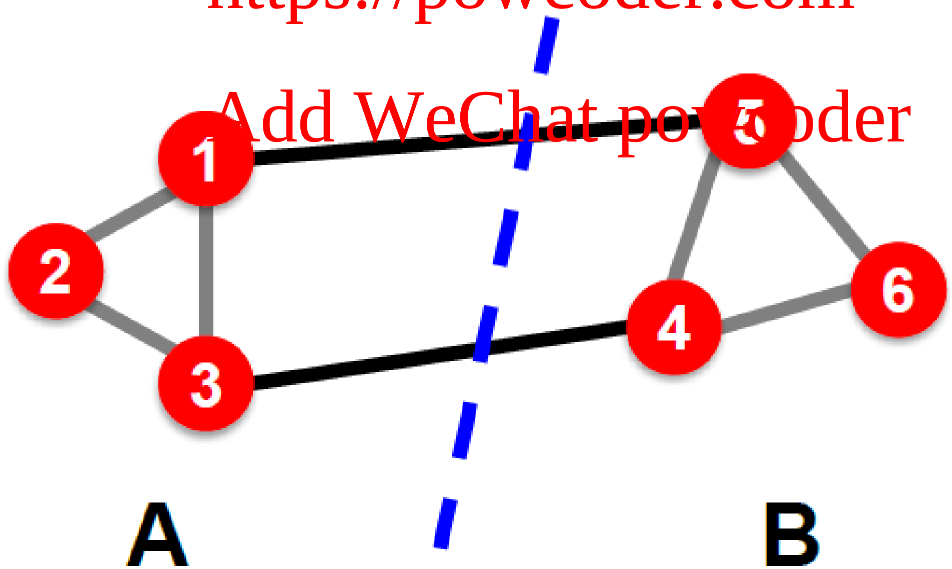


- **Questions:**
  - How can we define a "good" partition of G?
  - How can we efficiently identify such a partition?

Graph Partitioning

- **What makes a good partition?**
  - Maximize the number of within-group connections
  - Minimize the number of between-group connections

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Graph Cuts

- **Express partitioning objectives as a function of the "edge cut" of the partition**

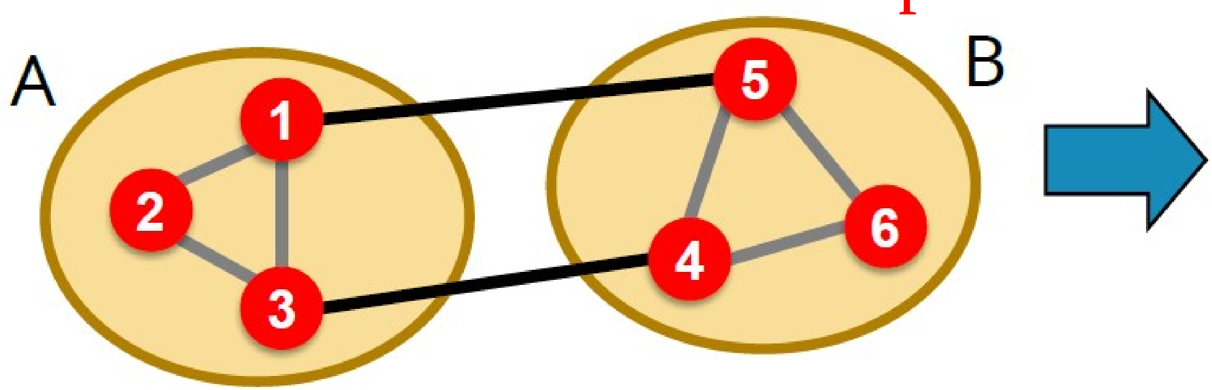
Assignment Project Exam Help

- **Cut:** Set of edges with only one vertex in a group:

https://powcoder.com

$$cut(A,B) = \sum_{i \in A, j \in B} w_{ij}$$

Add WeChat powcoder



$$cut(A,B) = 2$$

Graph Cut Criterion

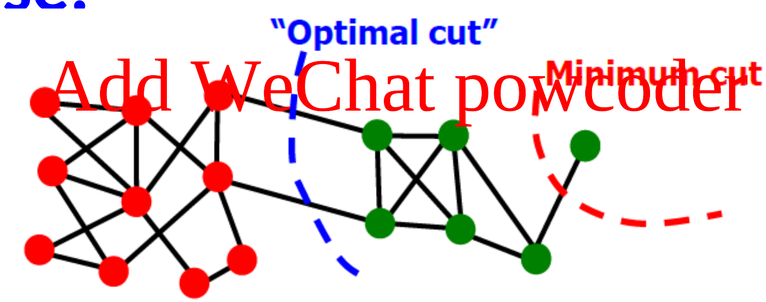
- **Criterion: Minimum-cut**
  - Minimize weight of connections between groups

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

- **Degenerate case:**



- **Problem:**
  - Only considers external cluster connections
  - Does not consider internal cluster connectivity

Graph Cut Criterion

- **Criterion: Normalized-cut** [Shi-Malik, '97]
  - Connectivity between groups relative to the density of each group

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

  - vol(A): total weight of the edges with at least one endpoint in A
- **Why use this criterion?**
  - Produces more balanced partitions
- **How do we efficiently find a good partition?**
  - **Problem:** Computing optimal cut is NP-hard

Spectral Graph Partitioning

- ***A***: adjacency matrix of undirected **G**
  - A_{ij} **=1** if (i,j) is an edge, else **0**

Assignment Project Exam Help

- ***x*** is a vector in $\Re^n$ with components $(x_1,\ldots,x_n)$. Think of it as a label/value of each node of G

https://powcoder.com

- **What is the meaning of *Ax*?**

Add WeChat powcoder

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad y_i = \sum_{j=1}^{n} A_{ij} x_j = \sum_{(i,j)\in E} x_j$$

- **Entry y_i is a sum of labels x_j of neighbors of *i***

What's the Meaning of Ax?

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- ***jth* coordinate of *Ax*:**
  - Sum of the ***x***-values of neighbors of ***j***
  - Make this a new value at node ***j***

Assignment Project Exam Help

$$\boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{\lambda} \cdot \boldsymbol{x}$$

- **Spectral Graph Theory:**

https://powcoder.com

  - Analyze the "spectrum" of matrix representing $\boldsymbol{G}$

Add WeChat powcoder

  - **Spectrum:** Eigenvectors $\boldsymbol{x}_i$ of a graph, ordered by the magnitude (strength) of their corresponding eigenvalues $\boldsymbol{\lambda}_i$:

$$\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$$

$$\lambda_1 \le \lambda_2 \le \dots \le \lambda_n$$

Example: d-Regular Graph

- Suppose all nodes in G have degree d and G is connected
  - **What are some eigenvalues/vectors of G?**
  - $A \cdot x = \lambda \cdot x$ What is λ? What x?
    - **Let's try:** $x = (1,1,\ldots,1)$
    - **Then:** $A \cdot x = d,d,\ldots,d = \lambda \cdot x$. **So:** $\lambda = d$
  - **We found eigenpair of G:** $x = (1,1,\ldots,1)$, $\lambda = d$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Remember the meaning of $y = A \cdot x$:

$$y_i = \sum_{j=1}^{n} A_{ij} x_j = \sum_{(i,j) \in E} x_j$$

d is the Largest Eigenvalue of A

- **G** is **d**-regular connected, **A** is its adjacency matrix
- **Claim:**
  - **d** is largest eigenvalue of **A**.
  - **d** has multiplicity of **1** (there is only **1** eigenvector associated with eigenvalue **d**)
- **Proof: Why no eigenvalue $d' > d$?**
  - To obtain **d** we needed $x_i = x_j$ for every i,j
  - This means $x = c \cdot (1,1,\ldots,1)$ for some const. c
  - **Define:** S = nodes i with maximum possible value of x_i
  - Then consider some vector y which is not a multiple of vector $(1,\ldots,1)$. So not all nodes i (with labels y_i) are in S
  - Consider some node $j \in S$ and a neighbor $i \notin S$ then node j gets a value strictly less than d
  - So y is not eigenvector! And so d is the largest eigenvalue!

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Example: Graph on 2 Components

- What if G is not connected?
  - G has 2 components, each d-regular

Assignment Project Exam Help

- **What are some eigenvectors?**

https://powcoder.com

  - $\boldsymbol{x}$= Put all $\mathbf{1}$s on $\boldsymbol{A}$ and $\mathbf{0}$s on $\boldsymbol{B}$ or vice versa

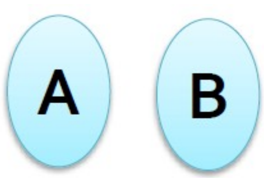
$$\boldsymbol{x}' = (\underbrace{\mathbf{1}, \dots, \mathbf{1}}_{|A|}, \underbrace{\mathbf{0}, \dots, \mathbf{0}}_{|B|}) \text{ then } \boldsymbol{A} \cdot \boldsymbol{x}' = (\boldsymbol{d}, \dots, \boldsymbol{d}, \mathbf{0}, \dots, \mathbf{0})$$

Add WeChat powcoder

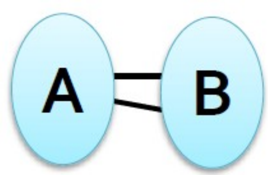
$$\boldsymbol{x}'' = (\mathbf{0}, \dots, \mathbf{0}, \mathbf{1}, \dots, \mathbf{1}) \text{ then } \boldsymbol{A} \cdot \boldsymbol{x}'' = (\mathbf{0}, \dots, \mathbf{0}, \boldsymbol{d}, \dots, \boldsymbol{d})$$

  - And so in both cases the corresponding

- A bit of intuition



$\lambda_n = \lambda_{n-1}$

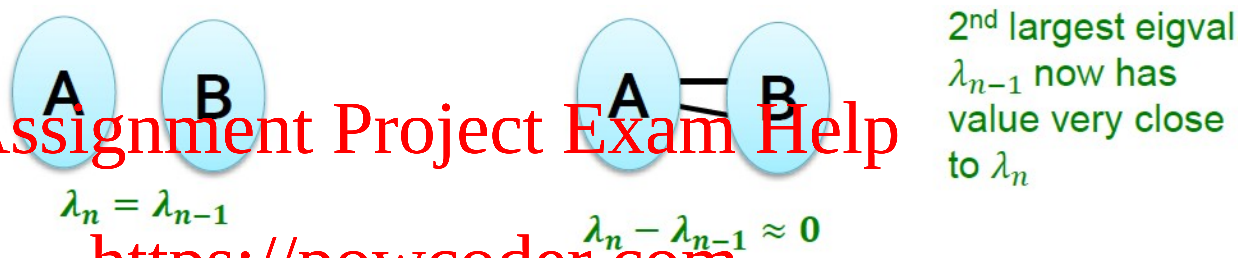


$\lambda_n - \lambda_{n-1} \approx \mathbf{0}$

2nd largest eigval. λ_{n-1} now has value very close to λ_n

More Intuition

- More Intuitions



Assignment Project Exam Help
https://powcoder.com

- If the graph is connected (right example) then we already know that $\boldsymbol{x}_n=(\mathbf{1},\ldots,\mathbf{1})$ is an eigenvector.

Add WeChat powcoder

  - Since eigenvectors are orthogonal then the components of $\boldsymbol{x}_{n-1}$ sum to **0**.
- So we can look at the eigenvector of the 2nd largest eigenvalue and declare nodes with positive label in **A** and negative label in **B**. **But there is still lots to sort out.**

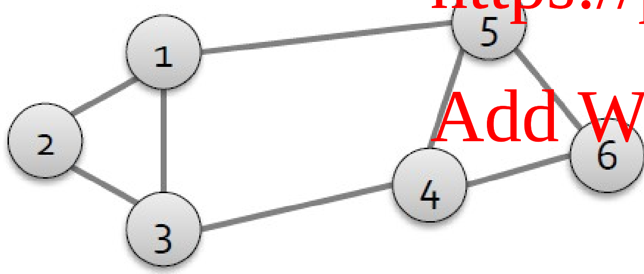
Matrix Representation

- **Adjacency matrix (*A*):**
  - ***n*n*** matrix
  - ***A=[a_{ij}], a_{ij}=1*** if edge between node ***i*** and ***j***

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder



	1	2	3	4	5	6
1	0	1	1	0	1	0
2	1	0	1	0	0	0
3	1	1	0	1	0	0
4	0	0	1	0	1	1
5	1	0	0	1	0	1
6	0	0	0	1	1	0

- **Important properties:**
  - Symmetric matrix
  - Eigenvectors are real and orthogonal

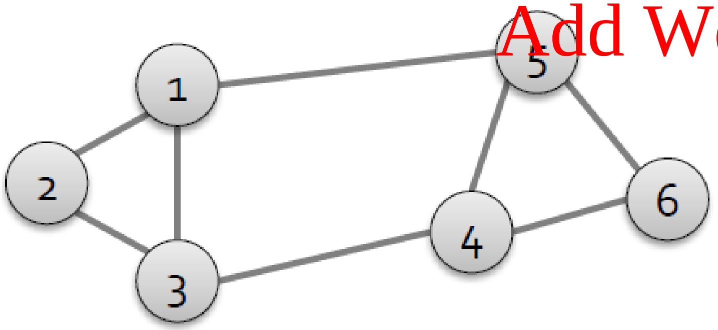
Matrix Representation

- **Degree matrix (D):**
  - $n*n$ diagonal matrix
  - $D=[d_{ii}]$, d_{ii} = degree of node i

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

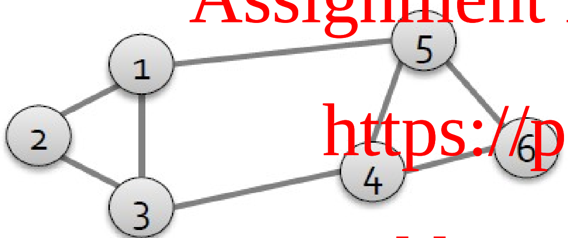


	1	2	3	4	5	6
1	3	0	0	0	0	0
2	0	2	0	0	0	0
3	0	0	3	0	0	0
4	0	0	0	3	0	0
5	0	0	0	0	3	0
6	0	0	0	0	0	2

Matrix Representation

- **Laplacian matrix (L):**
  - n^*n symmetric matrix

Assignment Project Exam Help



https://powcoder.com

	1	2	3	4	5	6
1	3	-1	-1	0	-1	0
2	-1	2	-1	0	0	0
3	-1	-1	3	-1	0	0
4	0	0	-1	3	-1	-1
5	-1	0	0	-1	3	-1
6	0	0	0	-1	-1	2

Add WeChat powcoder

$$L = D - A$$

- **What is trivial Eigenpair?**
  - $x=(1,\ldots,1)$ then $L \cdot x = 0$ and so $\lambda = \lambda_1 = 0$
- **Important properties:**
  - **Eigenvalues** are non-negative real numbers
  - **Eigenvectors** are real and orthogonal

Facts about Laplacian L

- **(a)** All eigenvalues are ≥ 0
- **(b)** for every x
- **(c)**

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

- That is, L is **positive semi-definite**

as Optimization Problem

- **Fact: For symmetric matrix *L*:**

Assignment Project Exam Help

https://powcoder.com

- **What is the meaning of min $x^{T} L x$ on *G*?**

Add WeChat powcoder

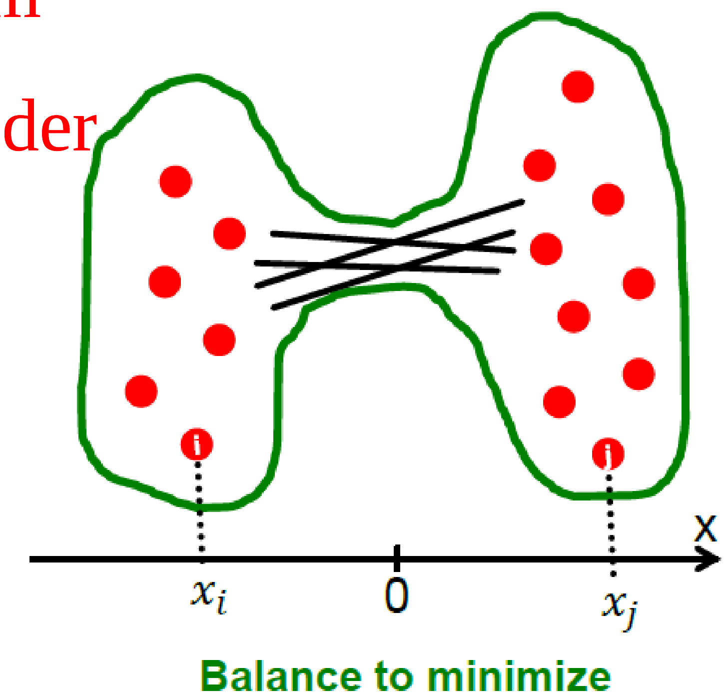
  - $x^{T} L x = \sum_{i,j=1}^{n} L_{ij} x_i x_j = \sum_{i,j=1}^{n} (D_{ij} - A_{ij}) x_i x_j$
  - $= \sum_i D_{ii} x_i^2 - \sum_{(i,j) \in E} 2 x_i x_j$
  - $= \sum_{(i,j) \in E} (x_i^2 + x_j^2 - 2 x_i x_j)$

as Optimization Problem

- **What else do we know about *x*?**
    - $\boldsymbol{x}$ is unit vector:
    - $\boldsymbol{x}$ is orthogonal to **1st** eigenvector $(\mathbf{1},\ldots,\mathbf{1})$ thus:

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- Remember



We want to assign values x_i to nodes i such that few edges cross 0. (we want x_i and x_j to subtract each other)

Finding Optimal Cut [Fielder'73]

- Back to finding the optimal cut
- Express partition (A,B) as a vector

Assignment Project Exam Help

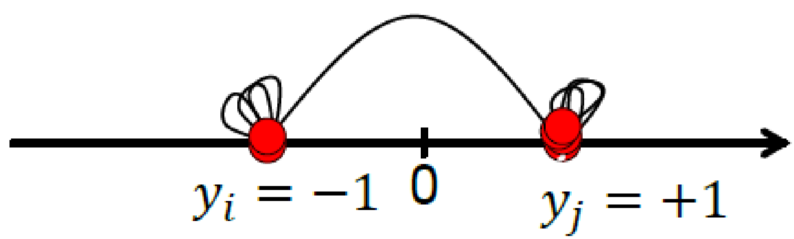
$$\boldsymbol{y_i} = \begin{cases} +\mathbf{1} & \boldsymbol{if\ i \in A} \\ -\mathbf{1} & \boldsymbol{if\ i \in B} \end{cases}$$

https://powcoder.com

Add WeChat powcoder

- We can minimize the cut of the partition by finding a non-trivial vector *x* that **minimizes**

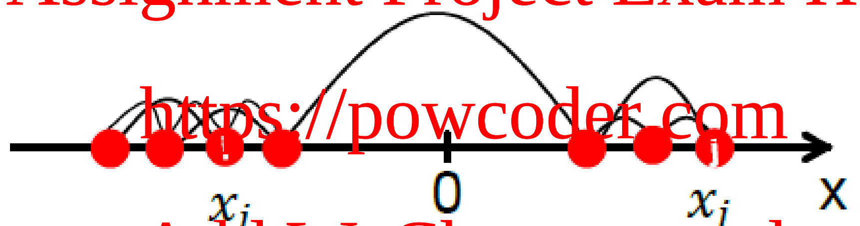
Can't solve exactly. Let's relax y and allow it to take any real value.



Rayleigh Theorem

Assignment Project Exam Help

https://powcoder.com



Add WeChat powcoder

- The minimum value of $\boldsymbol{f}(\boldsymbol{y})$ is given by the 2nd smallest eigenvalue λ_2 of the Laplacian matrix $\boldsymbol{L}$
- The optimal solution for $\boldsymbol{f}(\boldsymbol{y})$ is given by the corresponding eigenvector $\boldsymbol{x}$, referred as the **Fiedler vector**

Approx. Guarantee of Spectral (details)

- Suppose there is a partition of **G** into **A** and **B** where $A \leq |B|$, s.t. then .
Assignment Project Exam Help
  - This is the approximation guarantee of the spectral clustering.
  - It says the cut spectral is at most **2** away from the optimal one of score .

https://powcoder.com

Add WeChat powcoder

So Far..

- **How to define a "good" partition of a graph?**
  - Minimize a given graph cut criterion
- **How to efficiently identify such a partition?**
  - Approximate using information provided by the eigenvalues and eigenvectors of a graph
- **Spectral Clustering**

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Spectral Clustering Algorithms

- **Three basic stages:**
  1) **Pre-processing**
     - Construct a matrix representation of the graph
  2) **Decomposition**
     - Compute eigenvalues and eigenvectors of the matrix
     - Map each point to a lower-dimensional representation based on one or more eigenvectors
  3) **Grouping**
     - Assign points to two or more clusters, based on the new representation

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Spectral Partition Algorithms

1) Pre-processing:

- Build Laplacian matrix L of the graph

	1	2	3	4	5	6
1	3	-1	-1	0	-1	0
2	-1	2	-1	0	0	0
3	-1	-1	3	-1	0	0
4	0	0	-1	3	-1	-1
5	-1	0	0	-1	3	-1
6	0	0	0	-1	-1	2

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

2) Decomposition:

- Find eigenvalues λ and eigenvectors x of the matrix L

λ =

0.0
1.0
3.0
3.0
4.0
5.0

X =

0.4	0.3	-0.5	-0.2	-0.4	-0.5
0.4	0.6	0.4	-0.4	0.4	0.0
0.4	0.3	0.1	0.6	-0.4	0.5
0.4	-0.3	0.1	0.6	0.4	-0.5
0.4	-0.3	-0.5	-0.2	0.4	0.5
0.4	-0.6	0.4	-0.4	-0.4	0.0

- Map vertices to corresponding components of λ_2

1	0.3
2	0.6
3	0.3
4	-0.3
5	-0.3
6	-0.6

How do we now find the clusters?

Spectral Partition Algorithms

- **3) Grouping**
  - Sort components of reduced 1-dimensional vector
  - Identify clusters by splitting the sorted vector in two
- **How to choose a splitting point?**
  - Naïve approaches:
    - Split at **0** or median value
  - More expensive approaches:
    - Attempt to minimize normalized cut in 1-dimension (sweep over ordering of nodes induced by the eigenvector)

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

1	0.3
2	0.6
3	0.3
4	-0.3
5	-0.3
6	-0.6

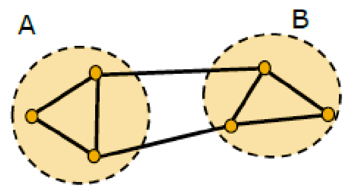
Split at 0:

Cluster A: Positive points

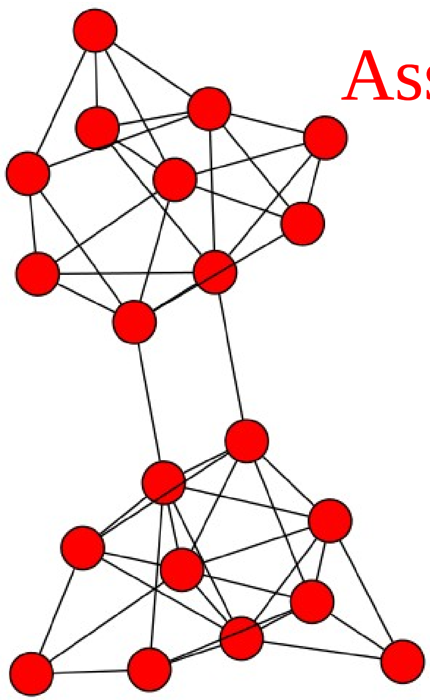
Cluster B: Negative points

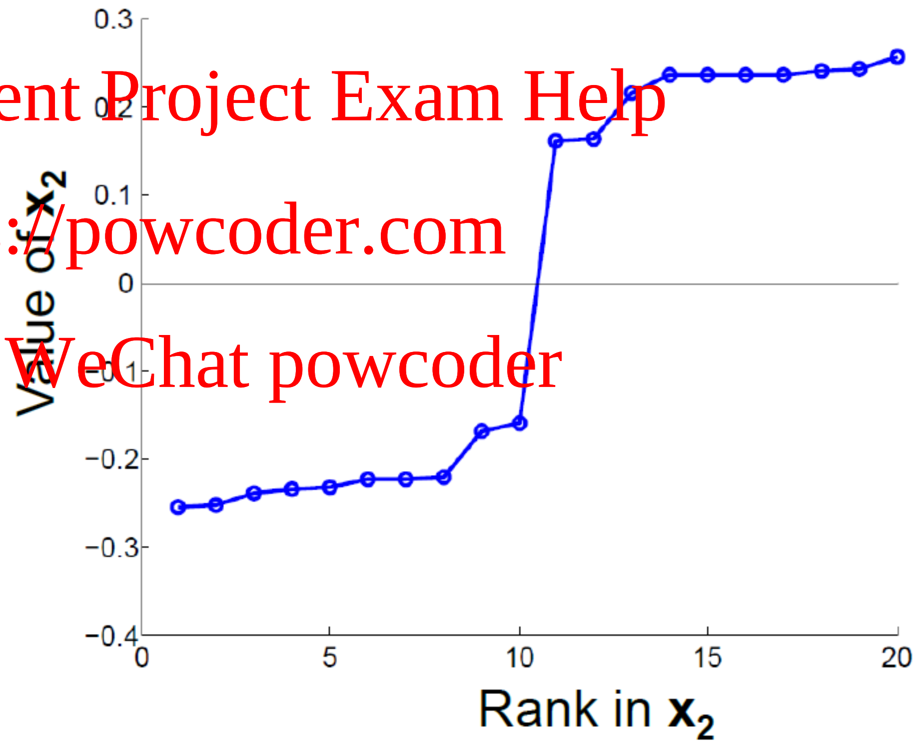
1	0.3
2	0.6
3	0.3

4	-0.3
5	-0.3
6	-0.6



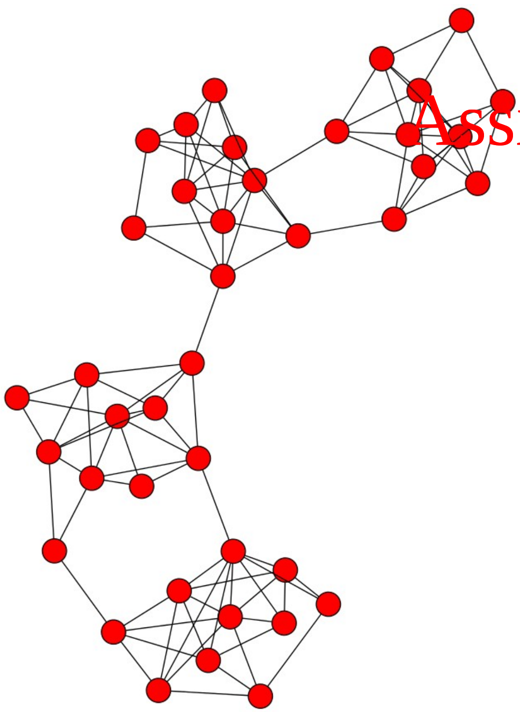
Example: Spectral Partitioning

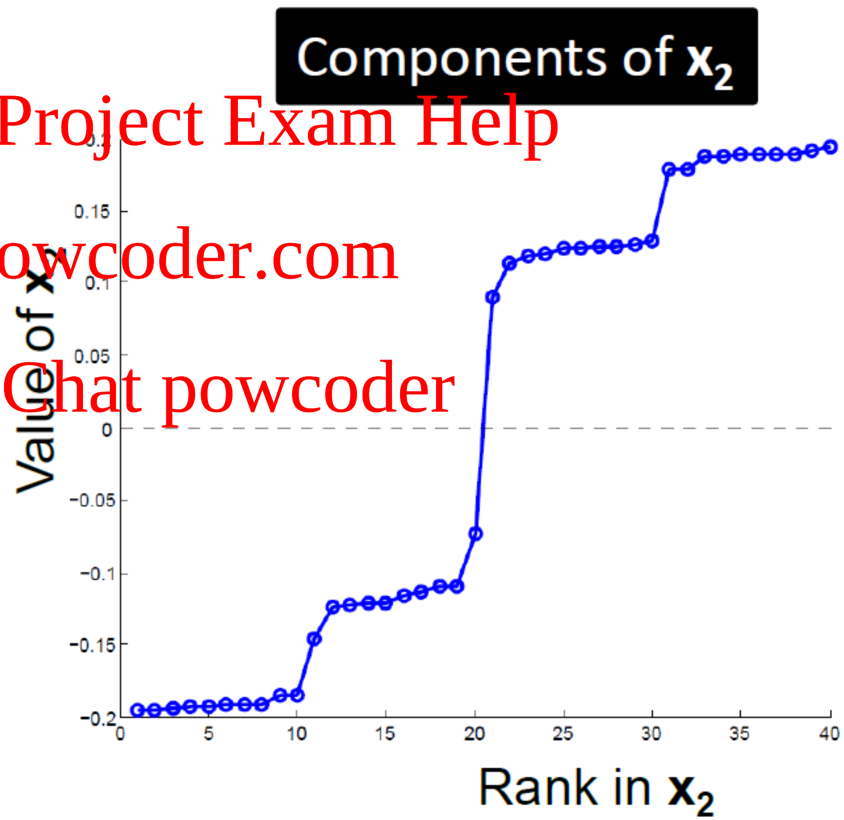




Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

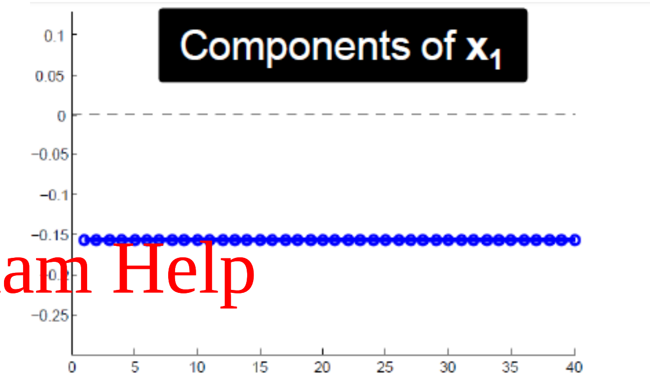
Example: Spectral Partitioning

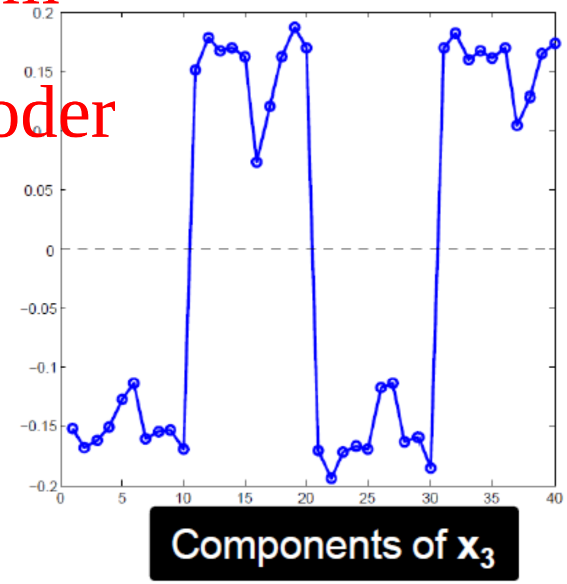




Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Example: Spectral Partitioning





Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Outline

- Introduction
- Social Networks as Graph
  - Basics of Graph Theory
  - Important properties of social networks as graphs
  - Analysis of a real-world social network
- Community detection algorithms
  - Girvan-Newman algorithm
  - Spectral clustering
- Conclusions

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Conclusions

- **How do we reason about networks?**
  - **Empirical:** Study network data to find organizational principles. How do we measure and quantify networks?

Assignment Project Exam Help

- **Mathematical models:**
  - Graph theory, statistical models
    - allow us to understand behaviors and distinguish surprising from expected phenomena

https://powcoder.com
Add WeChat powcoder

- **Algorithms**
  - for analyzing graphs
    - Hard computational challenges

Conclusions: This is Just a Beginning

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Observations	Models	Algorithms
Small diameter, Edge clustering	Small-world model, Erdös-Renyi model	Decentralized search
Scale-free	Preferential attachment, Copying model	PageRank, Hubs and authorities
Strength of weak ties, Core-periphery	Kronecker Graphs	Community detection: Girvan-Newman, Modularity
Densification power law, Shrinking diameters	Microscopic model of evolving networks	Link prediction, Supervised random walks
Patterns of signed edge creation	Structural balance, Theory of status	Models for predicting edge signs
Viral Marketing, Blogosphere, Memetracking	Independent cascade model, Game theoretic model	Influence maximization, Outbreak detection, LIM

One-Slide Takeaway

- Key concepts of social graphs
  - Undirected/Directed Graph/Adjacency Matrix
  - Shortest path/Diameter
  - Degree distribution/Path length/Clustering Coefficient
  - Real world graph
    - Degree distribution: high-skewed
    - Most path lengths are small
- Community detection algorithms
  - Purpose of community detection
  - Spectral clustering

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Reference

- Adamic, Lada A., and Eytan Adar. "Friends and neighbors on the web." Social networks 25.3 (2003): 211-230.
- Kossinets, Gueorgi, and Duncan J. Watts. "Empirical analysis of an evolving social network." Science 311.5757 (2006): 88-90. APA
- Liben-Nowell, David, et al. "Geographic routing in social networks." Proceedings of the National Academy of Sciences of the United States of America 102.33 (2005): 11623-11628.
- Leskovec, Jure, and Eric Horvitz. "Planetary-scale views on a large instant-messaging network." Proceedings of the 17th international conference on World Wide Web. ACM, 2008.
- J. Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE
- Trans. on Pattern Analysis and Machine Intelligence,” 22:8 (2000), pp. 888–905.
- S. Fortunato, “Community detection in graphs,” Physics Reports 486:3–5(2010), pp. 75–174.
- M. Girvan and M.E.J. Newman, “Community structure in social and bi-ological networks,” Proc. Natl. Acad. Sci. 99 (2002), pp. 7821–7826.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Reference

- U. von Luxburg, “A tutorial on spectral clustering,” Statistics and Computing bf17:4 (2007), 2007, pp. 395–416.
- J. Leskovec, K.J. Lang, A. Dasgupta, and M.W. Mahoney, “Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters,” http://arxiv.org/abs/0810.1355.
- L. Backstrom and J. Leskovec, “Supervised random walks: predicting and recommending links in social networks,” Proc. Fourth ACM Intl. Conf. on Web Search and Data Mining (2011), pp. 635–644.
- Predicting Positive and Negative Links in Online Social Networks by J. Leskovec, D. Huttenlocher, J. Kleinberg. ACM WWW International conference on World Wide Web (WWW), 2010.
- Learning to Discover Social Circles in Ego Networks by J. McAuley, J. Leskovec. Neural Information Processing Systems (NIPS), 2012.
- Defining and Evaluating Network Communities based on Ground-truth by J. Yang, J. Leskovec. IEEE International Conference On Data Mining (ICDM), 2012.
- The Life and Death of Online Groups: Predicting Group Growth and Longevity by S. Kairam, D. Wang, J. Leskovec. ACM International Conference on Web Search and Data Mining (WSDM), 2012

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

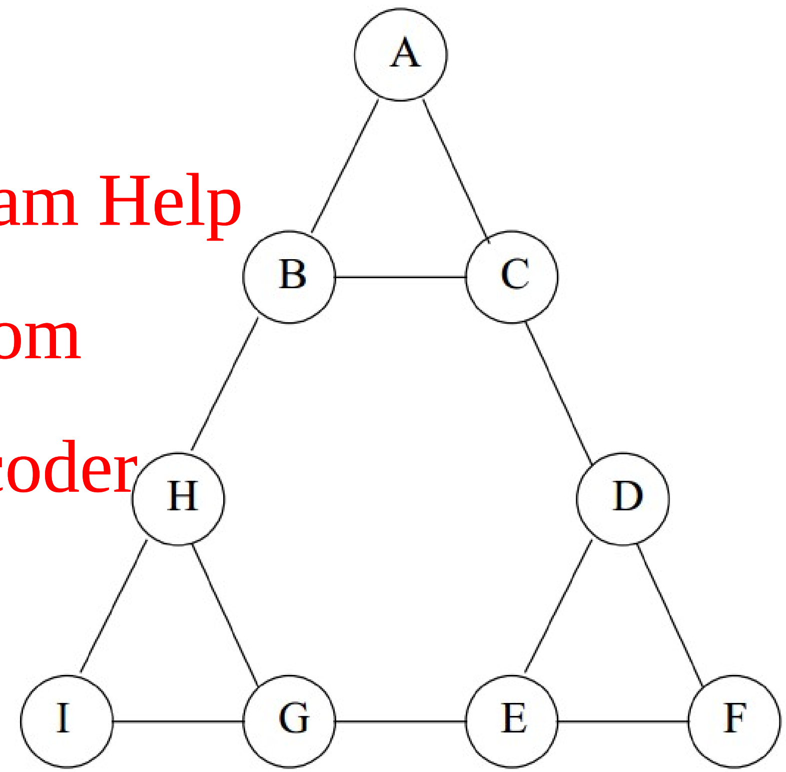
In-class Practice 1

- For the graph on the right, compute:
  - The adjacency matrix
  - The degree matrix
  - The Laplacian matrix

Assignment Project Exam Help

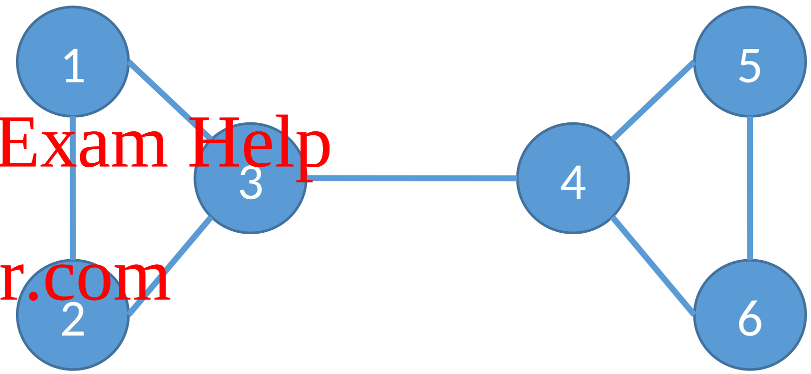
https://powcoder.com

Add WeChat powcoder



In-class Practice 2

- For the graph on the right, compute:
  - The shortest path between each pairs of nodes
  - The betweenness of each edge



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder