IT482 SNA TEAM ASSIGNMENT 1 - Information Network Analysis

Team Members

Subgroup 1

Jay Chavan [221IT020]
Nithin S [221IT085]

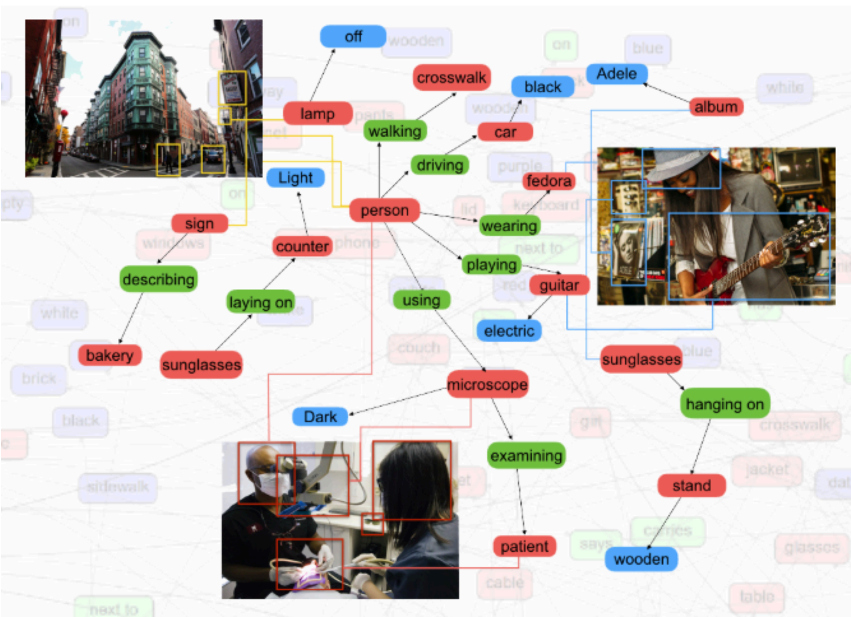
Subgroup 2

Sanketh N S [221IT060]
Vrushank T S [221IT083]

Design an information network analysis (INA) strategy for measuring the various network properties as listed below, supported by necessary visualizations.

Dataset Description :

VisualGenome



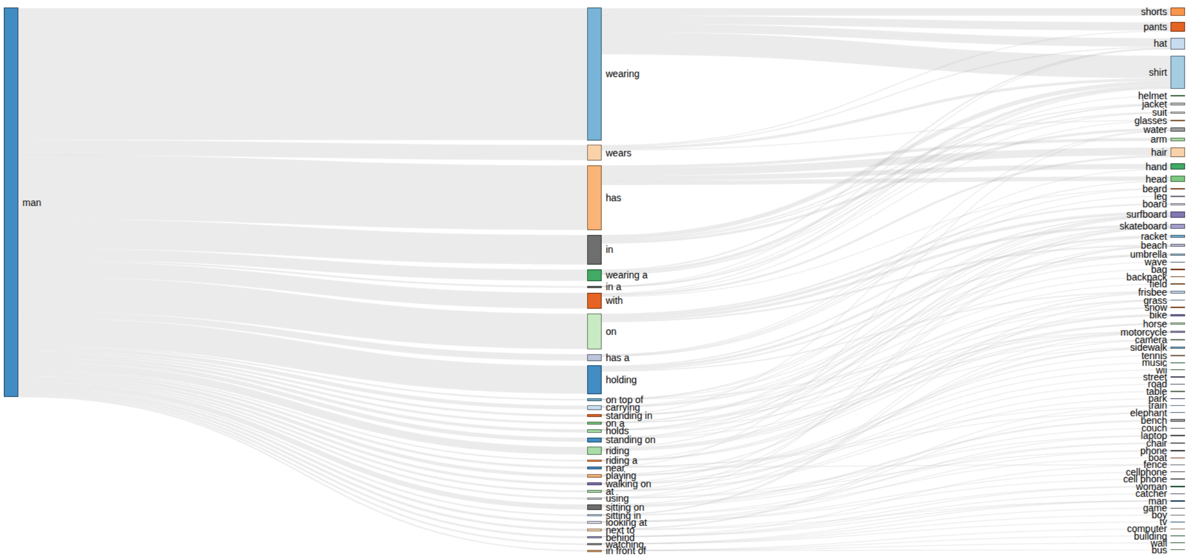
The Visual Genome dataset is a **large-scale visual knowledge base** designed to bridge the gap between **vision** and **language** by densely annotating images with rich textual information. It enables cognitive-level understanding of images, not just recognition

- **108,077 images**
- **5.4 million region descriptions**—sub-parts of images described in natural language.
- **1.7 million visual question–answer pairs**, enabling VQA tasks
- **3.8 million object instances** tagged with bounding boxes
- **2.8 million object attributes**, such as color, size, texture, etc.
- **2.3 million relationships** among objects (e.g., "person riding horse")
- All entities (objects, attributes, relationships, phrases from regions and QA pairs) are **canonicalized using WordNet synsets**, ensuring consistent semantic grounding

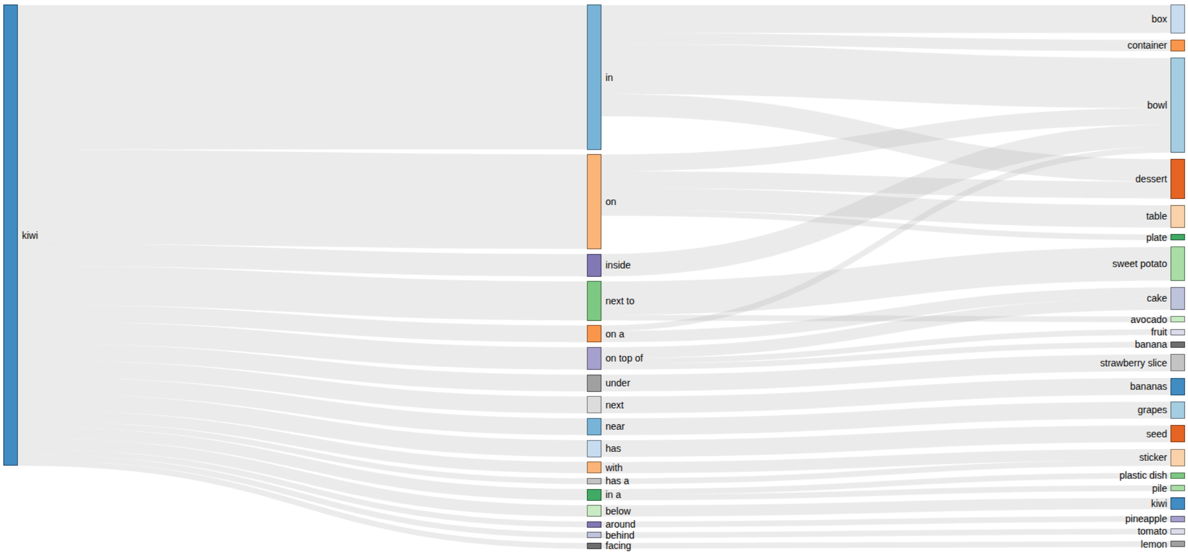
Dataset Visualization

relationships.json - has triplets {“man”, “has“, “water” }

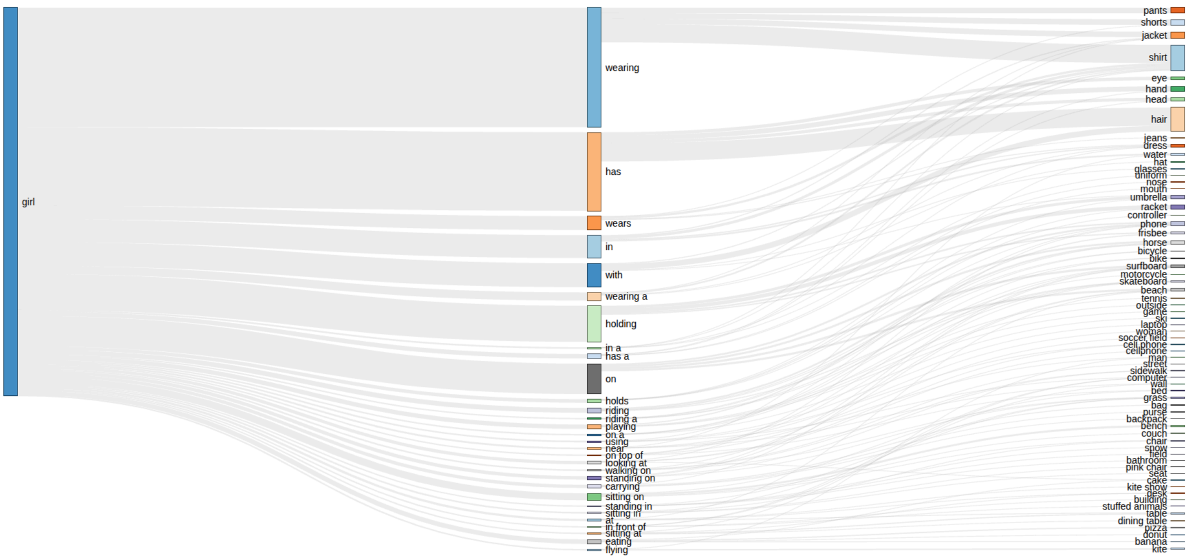
Man



Kiwi



Girl



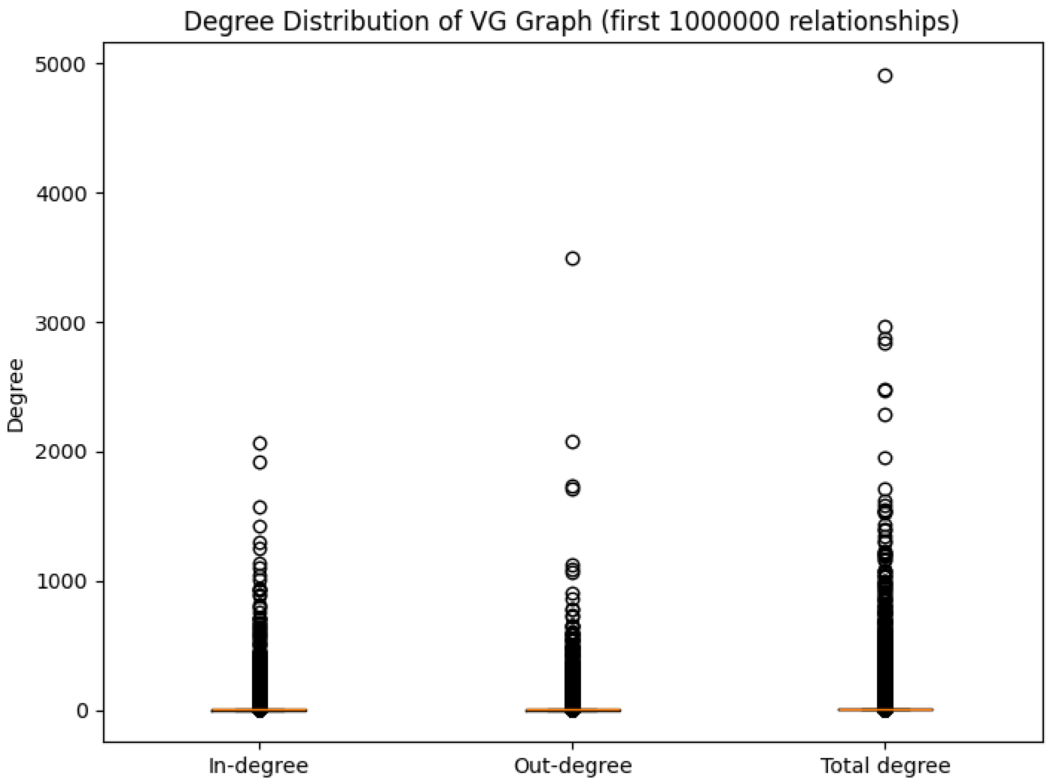
Subgroup 1 (Jay & Nithin) – **Local and Degree-centric Property Analysis of IN**

1. Average degree – *Compute average degree (and in/out-degree if directed), analyse the variation in average degree of nodes in the graph by generating box plots of computed node degree.*
2. Degree distribution – *Analyze and fit to theoretical models (e.g., Poisson, Power-law, Log-normal), report observations wrt interesting phenomena using log-log plots, and highlight outlier nodes that deviate strongly from the trend.*

3. Path length – *compute shortest, average, diameter; identify node pairs that achieve the diameter, and discuss whether the network shows “small-world” behaviour (average path length ~ log(N)).*
4. Geodesic path length – *Compute average geodesic path length, compare the value for entire graph vs. giant component (if applicable to your dataset)*
5. Clustering coefficient & average clustering coefficient – *Compare CC of low-degree nodes vs high-degree nodes, plot CC vs degree to illustrate differences between hubs and peripheral nodes.*

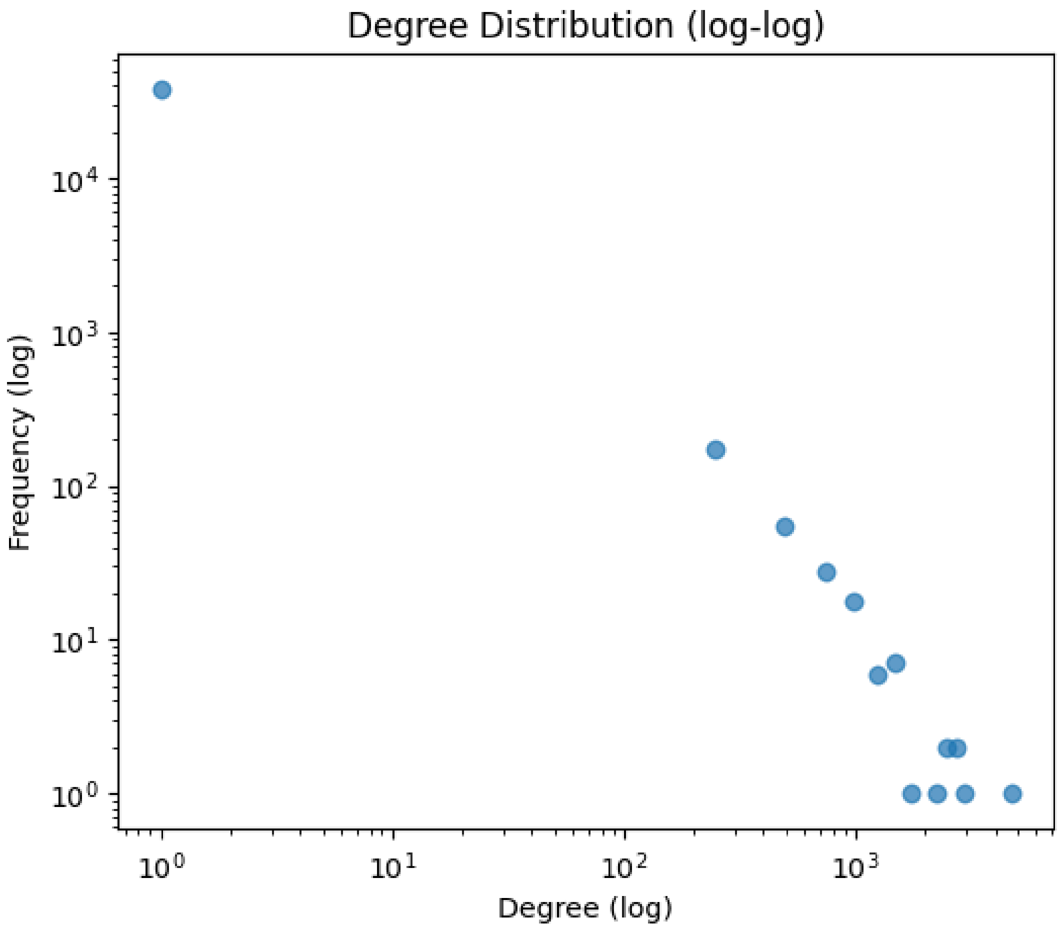
Solution:

Done only on first 1000K edges due to computation complexity

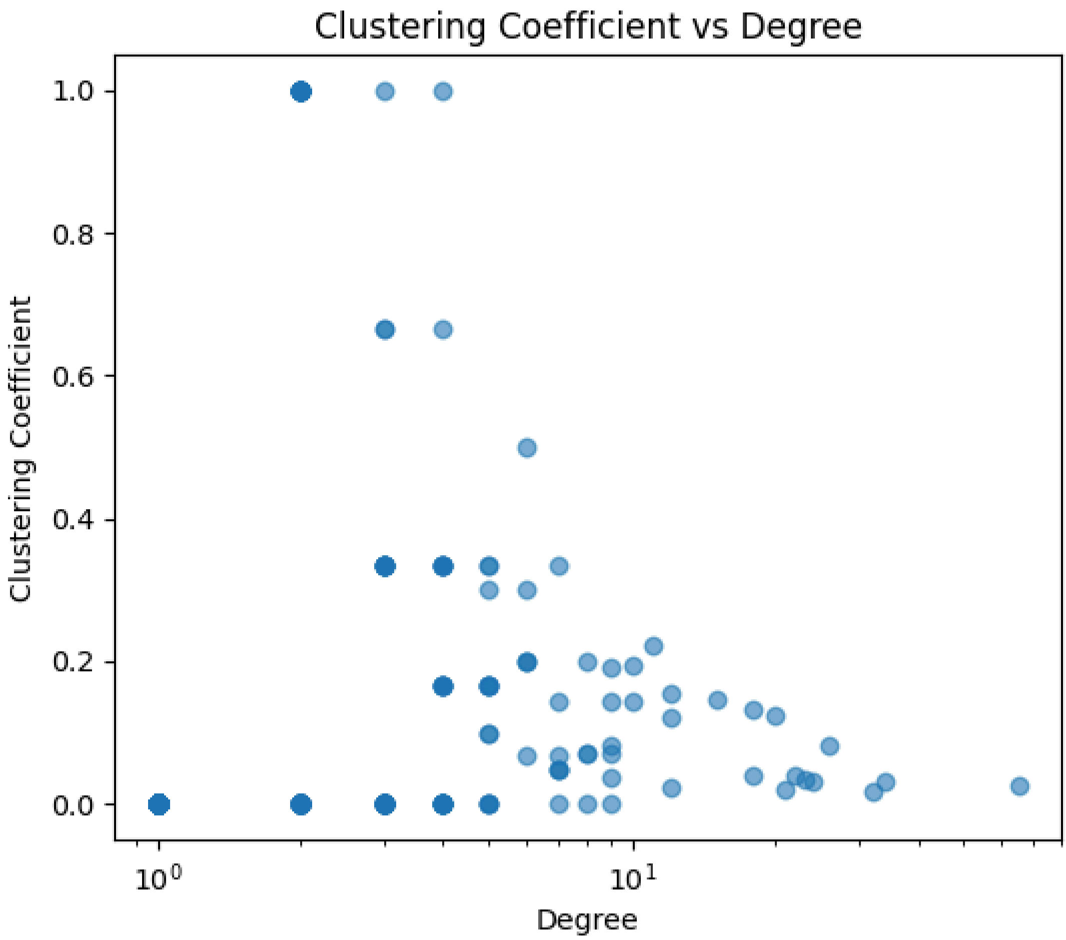



```
Average in-degree: 5.242467606806473
Average out-degree: 5.242467606806473
Average total degree: 10.484935213612946
```

```
Outlier nodes: ['sidewalk', 'man', 'shoes', 'car', 'sign', 'building', 'shirt', 'street', 'back', 'glasses', 'pants', 'jacket', 'bike', 'van', 'tree', 'trees', 'windows', 'backpack',
 'road', 'window', 'keyboard', 'monitor', 'desk', 'girl', 'computer', 'wall', 'hair', 'chain', 'pen', 'floor', 'bag', 'woman', 'picture', 'phone', 'handle', 'table', 'chair', 'carpet
', 'pillow', 'couch', 'teddy bear', 'animal', 'frame', 'door', 'lamp', 'books', 'shelf', 'railing', 'paper', 'room', 'apple', 'counter', 'lid', 'glass', 'top', 'fork', 'plate', 'bowl
', 'pan', 'food', 'container', 'cover', 'leg', 'bottle', 'logo', 'book', 'writing', 'board', 'laptop', 'label', 'hand', 'dish', 'boy', 'cup', 'front', 'guy', 'holder', 'lights', 'bab
y', 'tower', 'wheels', 'coat', 'wires', 'screen', 'box', 'sun', 'pot', 'plant', 'legs', 'spots', 'photo', 'tray', 'cord', 'people', 'painting', 'cap', 'red', 'jeans', 'male', 'pole',
 'light', 'tile', 'chairs', 'poster', 'cabinet', 'stand', 'can', 'helmet', 'pipe', 'wood', 'sofa', 'child', 'arm', 'neck', 'person', 'kid', 'train', 'grass', 'balcony', 'shadow', 'ca
rs', 'stop sign', 'ground', 'tire', 'advertisement', 'bicycle', 'stripes', 'poles', 'flowers', 'statue', 'buildings', 'men', 'reflection', 'seat', 'ceiling', 'vase', 'cloth', 'basket
', 'ocean', 'beach', 'hands', 'sky', 'boat', 'shorts', 'words', 'sand', 'water', 'house', 'women', 'luggage', 'tv', 'rug', 'remote', 'knife', 'oven', 'sink', 'rack', 'fireplace', 'wh
eel', 'base', 'kitchen', 'plants', 'mirror', 'flower', 'edge', 'background', 'pile', 'fridge', 'left', 'city', 'doorway', 'knob', 'television', 'white', 'brown', 'brick', 'line', 'co
rner', 'curb', 'vehicle', 'store', 'umbrella', 'street sign', 'group', 'tie', 'leaves', 'signs', 'flag', 'river', 'motorcycle', 'traffic light', 'head', 'windshield', 'fence', 'trash
 can', 'bus', 'letters', 'paint', 'design', 'lettering', 'stripe', 'truck', 'post', 'trunk', 'platform', 'stairs', 'wire', 'sunglasses', 'black', 'suitcase', 'green', 'right', 'metal
', 'cake', 'napkin', 'bar', 'curtain', 'jar', 'refrigerator', 'stove', 'bed', 'shoe', 'hat', 'rock', 'blanket', 'spoon', 'banana', 'tail', 'fruit', 'orange', 'trim', 'string', 'bread
', 'cheese', 'buttons', 'circle', 'roof', 'dirt', 'lines', 'blue', 'sticker', 'pitcher', 'strap', 'nose', 'ear', 'button', 'lady', 'row', 'rail', 'doors', 'hole', 'pavement', 'bench'
, 'wave', 'zebra', 'eyes', 'ramp', 'field', 'walkway', 'leaf', 'foot', 'ring', 'finger', 'eye', 'mouth', 'face', 'stem', 'object', 'track', 'branch', 'tag', 'feet', 'doughnut', 'donu
t', 'catcher', 'batter', 'cart', 'numbers', 'camera', 'bucket', 'bottom', 'side', 'microwave', 'stick', 'park', 'steps', 'meat', 'path', 'sandwich', 'sauce', 'bridge', 'mountain', 't
oilet', 'hill', 'display', 'name', 'letter', 'clock', 'bushes', 'ball', 'tracks', 'patch', 'case', 'gate', 'rocks', 'banner', 'bush', 'spot', 'ship', 'a', 'word', 'fire hydrant', 'bi
rd', 'stone', 'distance', 'collar', 'elephant', 'dog', 'hydrant', 'snow', 'couple', 'cat', 'scene', 'towel', 'surface', 'the', 'plane', 'uniform', 'crowd', 'concrete', 'air', 'cow',
'bathroom', 'part', 'toy', 'chest', 'bear', 'area', 'horse', 'number', 'airport', 'airplane', 'wing', 'engine', 'runway', 'mountains', 'yellow', 'body', 'net', 'giraffe', 'structure'
, 'pizza', 'jersey', 'player', 'tip', 'court', 'racket', 'section', 'lake', 'jet', 'glove', 'hot dog', 'sheep', 'carrot', 'bananas', 'rope', 'horses', 'tennis player', 'station', 'el
ephants', 'paw', 'zebras', 'animals', 'snowboard', 'kite', 'broccoli', 'tennis court', 'skateboard', 'giraffes', 'skier', 'surfboard', 'surfer', 'frisbee', 'baseball player', 'skateb
oarder']
```



Done on 1000 Edges due to computation complexity




```
Average path length: 4.012102902133114
Diameter: 10
log(N): 5.802118375377063
Example node pairs with diameter: [('shade', 'man'), ('shade', 'building'), ('shade', 'bottle'), ('shade', 'lights'), ('shade', 'screen')]
Average geodesic path length (giant component): 4.012102902133114
Average clustering coefficient: 0.10151197409394377
Avg CC (low-degree nodes): 0.0
Avg CC (high-degree nodes): 0.14068189343301182
```

Subgroup 2 (Sanketh & Vrushank) – **Global and Structural Property Analysis of IN**

1. Number of Strongly Connected Components (SCC) – *also report the size distribution of each SCC to highlight the variation in member node number.*
2. Number of Weakly Connected Components (WCC) – *analysis same as above (applicable for directed networks only)*
3. Giant component and coverage statistics -- *Identify giant components and report coverage (% of nodes and edges), compare properties of giant vs non-giant components (e.g., average degree, clustering).*
4. Giant component properties – *Compute shortest path, average path length, diameter, and Avg CC inside the giant component, compare values with those from*

the full graph, and discuss whether the giant component shows small-world features.

5. k-connectedness – *Analyze connectedness with respect to different values of k, Plot size of largest connected subgraph vs k, check if the IN is fragile or robust to increasing k.*

Solution:

Done only on first 1000 edges due to computation complexity

```
Number of SCCs: 329
SCC size distribution (top 10): [48, 3, 2, 2, 2, 2, 1, 1, 1, 1]
Number of WCCs: 26
WCC size distribution (top 10): [331, 4, 3, 3, 3, 2, 2, 2, 2, 2]

=== Giant Component ===
Giant nodes: 331/382 (86.65%)
Giant edges: 555/585 (94.87%)
Avg degree - full graph: 3.06282722513089 giant: 3.3534743202416917
Avg clustering - full graph: 0.08795932833794605 giant: 0.10151197409394377

Giant component avg path length: 4.012102902133114
Giant component diameter: 10
Giant component avg clustering: 0.10151197409394377
Random-graph expected clustering: 0.009685983704110593
Approx random path length: 4.795151363287065
Small-world features: LIKELY

=== K-core Decomposition ===
k=1: nodes=331, edges=512
k=2: nodes=138, edges=319
k=3: nodes=71, edges=192
k=4: nodes=22, edges=64
k=5: nodes=0, edges=0
```