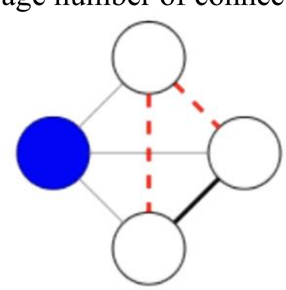
1. I have two examples

   a) In a city with many ride-sharing drivers and riders, the goal is to efficiently match drivers with riders while ensuring both parties are satisfied with their assignments. There are n ride-sharing drivers and n riders, where each driver has preferences for the riders they'd like to pick up, and each rider has preferences for the drivers they'd like to ride with. The objective is to create stable matches between drivers and riders such that neither party has an incentive to deviate from their assigned partner.

   b) One of the situations in urban settings I could think of is the struggle of young parents to choose the best schools for their kids based on the infrastructures and other amenities. There are tons of schools throughout the city and the responsibility of a parent is to prioritize the schools based on their needs. On the other hand, the number of seats/vacancies in the school is very limited and they will also admit the students based on their performance, the financial situation of their parents etc. This is a pure stable marriage problem where parents apply to schools based on their priorities and schools will either admit/reject based on the number of seats available/other factors. If both the priorities of the school and the parents match, then the child will be successfully enrolled in the school. Else the parents will have to choose from the next best school from their list and the cycle repeats.

2. If it's a complete graph, that is all the substations are connected to each other, then we can place the sensor in any one of the substations, since it's connected to all other substations, we can get the status of all the substations at once. We can add another backup sensor to any one of the other substations just in case the substation having the sensor fails. We can place the sensors to the substations(nodes) which have a high number of degrees such that they can cover a large number of neighboring substations. By strategically placing sensors on critical nodes, high-degree nodes, and nodes with high betweenness centrality, we can effectively monitor the state of the entire power network with the least number of sensors. This strategy leverages network science concepts to identify key nodes for monitoring and optimize sensor placement for efficient coverage and fault detection.

3. Clustering coefficient tells us how the neighbors of a node know each other, and it is represented by a formula $C_i = 2e_i/k_i(k_i-1)$ where e_i represents the number of edges between the neighbors of node i and k_i is the degree of node i. The redundancy of a vertex v on the other hand is defined as the average number of connections from a



neighbor of i to other neighbors of i.

In the above example, based on the clustering coefficient calculation, it is found that the clustering Coefficient $C_i = 2(1)/3(2) = 1/3$ the redundancy is calculated as: the first neighbor doesn't have any edges to the neighbor so it's 0. The second neighbor has one link to a neighbor and vice versa. So redundancy is calculated as (0+1+1)/3 = ⅔ Since there won't be any self-loops, the equation between local clustering coefficient and redundancy is Ri = (n-1) Ci where n is the number of neighboring nodes of node i.

4. a. We can do some changes in the map while maintaining the network structure.i. Silver Line: Silver Line starts from Wiehle-Reston East and ends at Largo Town Center.This line seems fine and no alteration is required.
   ii. Orange Line: This line starts at Vienna and ends at New Carrollton. It travels along the same path as the silver line from East Falls Church to Stadium-Armory transfer stations. So we can remove the orange line in between these transfer stations and continue as the silver line.
   iii. Blue Line: The blue line starts from Franconia-Springfield and ends at Largo Town Center. It covers the same routes as that of Silver line from Rosslyn transfer station and hence we can remove the Blue Line from Rosslyn
   iv. Yellow Line: This line starts at Huntington and ends at Greenbelt. This line seems fine and no alteration is required.
   v. Green Line: This line starts from Branch Ave and ends at Greenbelt. This line covers the same routes as that of Yellow line from LE'nfant Plaza and hence we can remove the green line from LE'nfant Plaza.
   vi. Red Line: This line starts from Shady Grove and ends at Glenmont. This line seems fine, and no alteration is required.
   Few possible lines we can add are Wiehle-Reston East to Glenmont, Shady Grove to Largo Town Center, Franconia-Springfield to Branch Ave, Vienna to Greenbelt. Our objective is forming a graph having high clustering and low diameter.

   In case if we are planning to add new lines, we need to ask the WMATA whether it is possible to lay tracks between the stations of interest (urban planning and construction), whether we can add new intermediary stations or transfer stations between the stations which can be utilized by the surrounding urban population?
   b. WMATA has decided to introduce 3 express trains. The best way to implement this is to check the number of tracks/rail line connections that each station has. It is great if we implement express trains to and from the transfer stations. Since the connections are more in transfer stations, this denotes the vast majority of urban population are using metro rails to travel to various other stations. Express trains from East Falls Church to Rosslyn, Gallery Park to Fort Totten, Metro Station to Stadium-Armory, and vice versa can be introduced. If the stations are located on the same track and if the distance is greater and if it has more degrees, we can introduce an express train. We can ask questions like what the population density around those stations is, based on the stats on an average how many people used to travel between the two stations of interest, what's the distance between the stations of interest and whether it is feasible to introduce express trains based on the condition of tracks.
5. Graphs over time: densification laws, shrinking diameters and possible explanations:

i. In this research paper, the authors are considering evolving graphs and are performing various observations on that. They are introducing 2 methods in this research paper namely Community Guided Attachment and The Forest Fire Model. The properties of interest in their studies are based on two fundamental parameters: the nodes degrees and the distances between pairs of nodes(diameter). This paper explains more about 2 main empirical observations namely Densification power laws and Shrinking diameters.

Densification power law states that the network is becoming denser with the average number of degrees of node increasing over time. Moreover, the densification follows a power-law pattern.

The Shrinking diameter states that the diameter shrinks as the network becomes denser over time. This is surprisingly opposite to the initial assumption that the diameter increases when more nodes are introduced at a later point of time.

The authors introduce a densification power law equation: $e(t) \propto n(t)^a$ where e(t) and n(t) denote the number of edges and nodes of the graph at time t, and a is an exponent that generally lies strictly between 1 and 2. Here exponent a = 1 corresponds to constant average degree over time, while a = 2 corresponds to an extremely dense graph where each node has, on average, edges to a constant fraction of all nodes.

The authors discuss more about related works referencing various other research and authors. Most of the related works are based on static single or multiple snapshots of a graph and on the assumption that the nodes of the graph will have constant degree throughout the time.

The authors then list out their observations using datasets from four different sources each having information about the time when each node was added to the network over a period of several years thus enabling the construction of a snapshot at any desired point in time. For each of datasets, the authors find a version of the densification power law from the equation $e(t) \propto n(t)^a$; the exponent a differs across datasets but remains remarkably stable over time within each dataset. They also find that the effective diameter decreases in all the datasets considered. The datasets used by the authors for observation consist of two citation graphs for different areas in the physics literature, a citation graph for U.S. patents, a graph of the Internet, and five bipartite affiliation graphs of authors with papers they authored. Overall, they are considering 9 different datasets from 4 different sources. The 4 different graphs used in the observation are ArXiv citation graph, Patents citation graph, Autonomous systems graph, and Affiliation graphs.

The ArXiv citation graph covers the citation of 29,555 papers with 352,807 edges. This data covers papers in the period from January 1993 to April 2003 (124 months). The graph observation shows us that the slope a= 1.68. This tells us that since a>1, the graph is becoming denser as the average node degree is not constant (ie a<>1).

Similarly, the Patent citation graph covers 3,923,922 patents with 6,522,438 citations over a span of 37 years. The result is almost similar to the ArXiv citation graph with a slope of exponential constant a = 1.66 indicating that the graph network is becoming denser over a period of time. Finally the Autonomous systems graph and Affiliation graphs follow the same trend proving the densification power law with exponential constant a to be 1.18 and 1.15 respectively.

The authors plot the diameter changes over time for the 4 graphs considered and it is found out that the diameters shrink over time. The authors performed experiments to account for possible sampling problems, the effect of disconnected components, the effect of the "missing past", and the dynamics of the emergence of the giant component. To resolve possible sampling problems, the authors applied the Approximate Neighborhood Function (ANF) approach. For Disconnected components, the authors calculated the effective diameter of one single giant component of the graph and the effective diameter of the entire graph and it turned out to be the same. For the Missing part problem, the authors plot three different plots: Diameter of entire graph, Post-t0 subgraph and Post-t0 subgraph, no past.It is found out that all three curves lie close to one another indicating that missing part does not have anything to do with the shrinking of diameters. In Emergence of a giant graph, The authors see that within a few years the giant component accounts for almost all the nodes in the graph. The effective diameter, however, continues to steadily decrease beyond this point. This indicates that the decrease is happening in a "mature" graph, and not because many small disconnected components are being rapidly glued together.

In the Proposed System Section, the authors explain about their 2 proposed methods namely Community Guided Attachment and The Forest Fire Model.

In community guided attachment, the authors begin by searching for self-similar, recursive structures. The authors introduce 2 structures. Basic Version of the model and Dynamic Community Guided Attachment.

In the Basic Version of the model, the authors represent the recursive structure of communities-within-communities as a tree Γ, of height H. They construct nodes V in the graph with the leaves of the tree and the height h(u,v) is considered to be the height of their least common ancestor. Various formulas are derived on several conditions. The Dynamic Community Guided attachment is similar to the basic except that the graphs have internal nodes. Here we consider the length d(u,w) as the distance from u to the height of their least common ancestor and from the least common ancestor to w and hence d(u,w) = 2h(u,w).

The next method is The Forest Fire Model which captures all properties of network datasets. In this model a new node v will choose its ambassador node w uniformly at random, and forms a link to w. Node v selects x links incident to w, choosing from among both out-links and in-links, but selecting in-links with probability r times less than out-links. This method is done recursively and nodes cannot be visited a second time, preventing the construction from cycling. The authors introduced an Extension to the Forest Fire Model by introducing the terms orphans and Multiple ambassadors. Orphans are isolated nodes without any links. The authors conclude their research paper by talking about Densification power law, Shrinking of effective Diameters, Forest Fire model and Community Guided Attachment model

ii. The author in this paper put forth an example of CS grad students meeting old CS Students who introduce them to other CS/Non CS students. In terms of Social Network, say facebook as an example, We will initially start with 0 friends at time t=0. But we can find ourselves having more facebook friends when we are active in the network at a given time t. This denotes that new friends(nodes) in the graph are introduced and new links(friend request/connections) are formed and our networks become denser and denser. We can make use of the Forest Fire Model as an example in our facebook example, where we will choose a friend as an ambassador and we will visit their posts and tags and give friend requests to their friends(friends of friends) and so on. Eventually the diameter shrinks as the number of edges increases connecting various nodes/friends and in turn our facebook network becomes dense.

6. Urban spatial order: street network orientation, configuration, and entropy:
   i.This research paper by Geoff Boeing illustrates the urban spatial orders and entropy of 100 large cities across North America, South America, Europe, Africa, Asia, and Oceania. The intersections and deadends are considered as nodes and the street segments which link them are considered as edges. In this research paper, the author chose OpenStreetMap as a data source which is a collaborative worldwide mapping project that includes streets, buildings, amenities, and other spatial features. The author uses OSMnx software to download the street network within each city boundary from OpenStreetMap and then calculate several indicators. The autor calculates the street network's edges' individual compass bearings with OSMnx using two different methods. In the first method, the bearing of edge euv equals the compass heading from u to v and its reciprocal is captured. This method covers the orientation of the street but ignores the mid-block curvature. The second method covers it by weighting each edge's bearing by length to adjust for extremely short edges in these curve-approximations. In both the methods self-looping edges are ignored. Once all the edge compass bearings are calculated, it is divided into 36 equal-sized bins in which each bin represents 10o . To avoid extreme bin-edge effects around common values like 0° and 90°, the author shifts each bin by − 5° so that these values sit at the centers of their bins rather than at their edges. The Shannon entropy is calculated once the bearings are binned.
   For each city's graph, the entropy of the unweighted/simplified street orientations, Ho, is calculated as:

$$\mathbf{H_o = - \Sigma_{ni=1} P(o_i)log_eP(o_i)}$$

   where n represents the total number of bins, i represents the index of the bins, P(oi) represents the portion of the segment that falls in the ith bin.

The author calculates the entropy of weighted orientation Hw as well.

$$H_w = -\sum_{i=1}^{n} P(w_i)\log_e P(w_i)$$

Here Hw is biased by the city's shape (due to length-weighting), Ho is not.

The value of H is in dimensionless units called "nats," or the natural unit of information. The maximum value of entropy Hmax is calculated to be 3.584 nats. This represents a uniform distribution of street bearings across all the bins. If all the street bearings fall on a single bin, the entropy will be 0. But in the real world this isn't plausible as a more plausible minimum would instead be an idealized city grid with all streets in four equal proportions(north-south-east-west). This perfect grid entropy, Hg , would equal 1.386 nats.

Using Hg , Hmax and Ho we can calculate a normalized measure of orientation-order, φ, to indicate where a city stands on a linear spectrum from completely disordered/uniform to perfectly ordered/grid-like as:

$$\Phi = 1 - ((H_o - H_g) / (H_{max} - H_g))^2$$

The value of Φ ranges from 0 to 1. If the value of Φ is more, then it denotes that the city is more of a grid-like with minimum entropy and if the value of Φ is less then it denotes that the streets are uniformly distributed in every direction and have more entropy. The author calculates other indicators like each city's median street segment length. ĩ, average node degree k (i.e., how many edges are incident to the nodes on average), proportion of nodes that are dead-ends Pde, and proportion of nodes that are four-way intersections P4w. Finally, the author calculates each city street network's average circuity, ς as:

$$\varsigma = (L_{net}/L_{gc})$$

where Lnet represents the sum of all edge lengths in the graph and Lgc represents the sum of all great-circle distances between all pairs of adjacent nodes. ς represents how much more circuitous a city's street network is than it would be if all its edges were straight-line paths between nodes. Finally, the author considers all the key indicators calculated (k, φ, ĩ, and ς, Ho, Hw, Pde, P4w) in order to visualize the urban spatial orders of the 100 cities.

The author represents the derived data in a tabular form. Ho and Hw are strongly correlated. Based on the readings, it is found that Three American cities (Chicago (φ of 0.90), Miami (φ of 0.811), and Minneapolis (φ of 0.749)) have the lowest orientation entropies of all the cities studied, indicating that their street networks are the most ordered. It is found that most of the American cities have low orientation entropies and are more grid-like. Ironically one of the US city Charlotte (φ of 0.002) has high entropy and the streets are distributed in all directions.

In addition to this observation, Venice, Mogadishu, Helsinki, Jerusalem, and Casablanca have the shortest median street segment lengths while Kiev, Moscow, Pyongyang, Beijing, and Shanghai have the longest. Buenos Aires, Detroit, and Chicago have the least circuitous networks, while Caracas, Hong Kong, and Sarajevo have the most circuitous networks. Helsinki and Bangkok have the lowest average node degrees, each with fewer than 2.4 streets per node. Buenos Aires and Manhattan have the greatest average node degrees, both over 3.5 streets per node. Buenos Aires and Manhattan similarly have the largest proportions of four-way intersections and the smallest proportions of dead-end nodes.

To illustrate the geography of these order/entropy trends, the author maps the 100 study sites by φ terciles. Most of the sites in the US and Canada fall in the highest tercile. Most of the sites across the Middle East and South Asia fall in the middle tercile.

The author further uses various polar histograms, dendrograms, scatterplots in order to visualize the urban spatial ordering stating that the cities with low entropy and high φ value tend to be grid like and cities with high entropies are uniformly distributed in all directions.

ii. This approach seems to be a different way of representing street networks using compass bearings of streets and representing the observations in the form of 36 equally sized bins. It is really fascinating to find out that the streets in major US cities are linked together in the form of a grid which tends toward north-south-east-west. The nodes here are intersections and dead ends and the edges represent the segments connecting them. From this research paper, we learnt more about the urban spatial ordering of major cities throughout the world, the role of entropy in order to determine randomness and how uniformly the streets are distributed throughout the city.