

Lesson 10

Social Network Graph, Centralities, PageRank and KNN Analysis

•

Social Networking

- The grouping of individuals into specific groups, like small rural communities or some other neighbourhoods based on a requirement.

Social Network

- Is a social structure made of individuals (or organizations) called “nodes,” which are tied (connected) by one or more specific types of inter-dependency, such as friendship, kinship, financial exchange, dislike or relationships of beliefs, knowledge or prestige. (*Wikipedia*)

Social Network as Graphs

- Number of metrics for analysis
- Network topological analysis tools
- The degree, closeness, betweenness, egonet, K-neighbourhood, top-K shortest paths, PageRank, clustering, SimRank, connected components,

Social Network as Graphs

- K-cores, triangle count, graph matches and clustering coefficient.
- Collaborative filtering by bipartite weighted graph matching

Social Network Graph Analytics

- Compute degree centralities, degree distribution, separation of degrees, betweenness centralities, closeness centralities, neighbourhoods, PageRank, shortest path, Breadth First Search, minimum spanning tree (forest), connected components, spectral clustering and cluster coefficient.

Important Metrics— Centralities

- Degree (centrality)
- Closeness (centrality)
- Betweenness (centrality)
- Eigen vector (centrality)
- Eigen vector elements, such as status, rank and other properties

Social graph-network analytics

- Discovers the degree of interactions,
- Closeness,
- Betweenness,
- Ranks,
- Probabilities,
- Beliefs, and
- potentials.

Closeness and Sparseness

- Enables detection of abnormality in persons
- Abnormality found from properties of vertices and edges in network graph
- Analysis enables summarization and find attributes for *anomaly*.

Social Network Characteristics

1. Three-step neighbourhoods show positive correlation between a person and high performance

Betweenness between vertices and bridges between numbers of structures are not helpful to the organization.

Too many strong links of a person may have a negative correlation with

the performance.

Social Network Characteristics

1. Three-step neighbourhoods show positive correlation between a person and high performance

Betweenness between vertices and bridges between numbers of structures are not helpful to the organization.

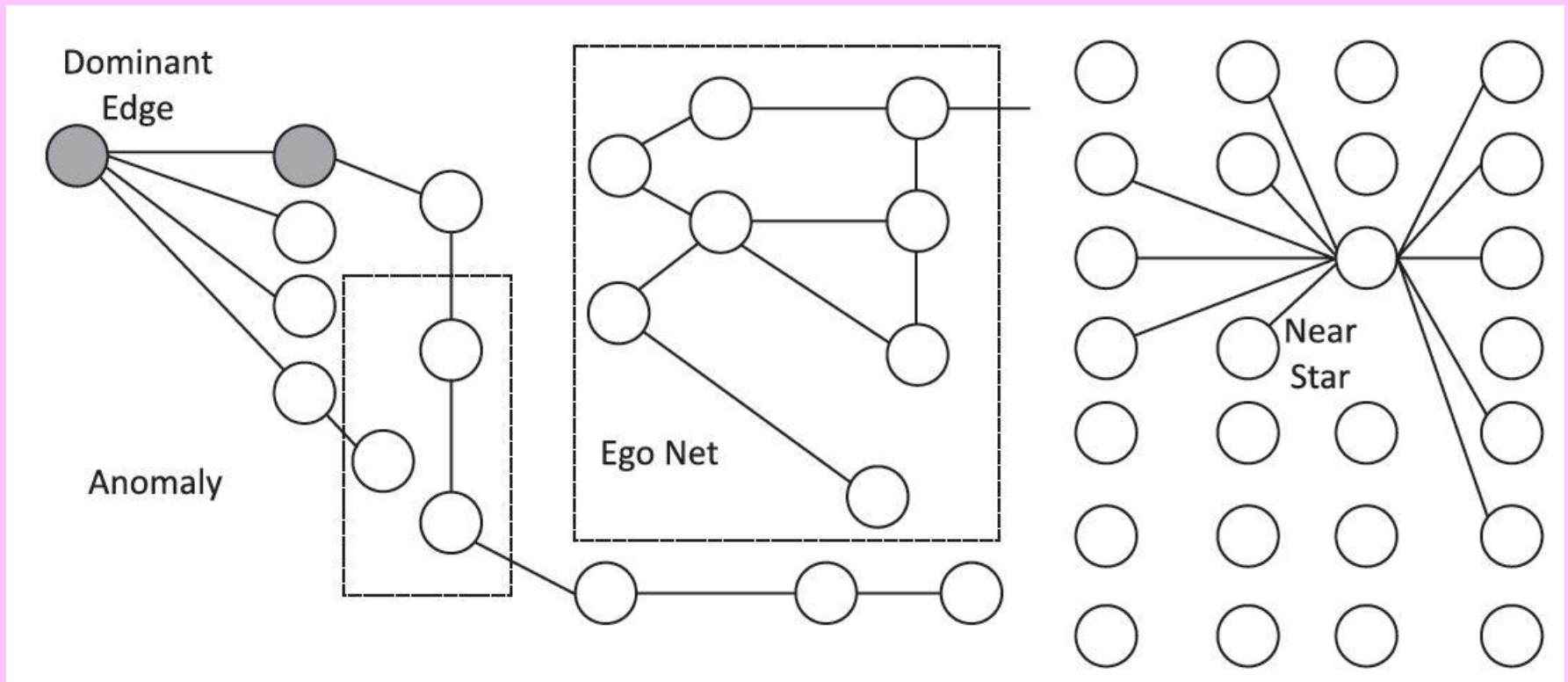
Characteristics

- Too many strong links of a person may have a negative correlation with the performance
- Social network of a person shows high performance outcome when the network exhibits structural diversity.

Characteristics

- Person with a social network with an abundant number of structural holes exhibits higher performance. This is because having diverse relations help an organization.

Figure 9.12 Discovering anomaly, ego-net and spam (using near star) from the analysis



Degree of a graph vertex

- Degree means the total number of edges linked to that
- In-degree the number of in-edges
- Out-degree the number of out-edges
- Degree distribution function (Section 6.2.5)

Graph vertex closeness $c_C(v)$

- A way of defining the centrality of a vertex in reference to other vertices
- Sum is the overall vertices connected to other vertices u . The u is a subset of vertices in set V .
- The centrality (closeness index), c is function of reciprocal of distances of vertices.

Effective Closeness

- Effective closeness $C_{ec}(v)$ can also be analyzed
- Use approximate average distance from v to all other vertices in place of the shortest paths
- C_{ec} reduces run time for cases with a large number of edges and near linear scalability in computations

Graph vertices Betweenness

- Means the number of times a vertex exists between the shortest path and the extent to which a vertex is located ‘between’ other pairs of vertices.
Betweenness $c_B(v)$ requires calculating the lengths of shortest paths among all pairs of vertices and computations of the summation for each pairing vertex in V

PageRank

- PageRank is a metric for the importance of each vertex in a graph, assuming an edge from v_1 to v_2 represents endorsement of importance of v_2 by v_1 by connecting, following, interacting, opting for relationship, sharing belief or some other means.

Structure Diversity

- Means that social graph has access to diverse sub-graphs (knowledge).

K-core

- Is a sub-graph in a graph network structure
- Graph Vertex K^{th} neighbourhood is number of 1st neighbour vertices, 2nd neighbour vertices and so on to a querying vertex that are correlated, linked, and have weighted correlations or the associations.

K-nearest neighbourhood (KNN)

- Finds K-similar objects, items, or entities, which are nearest neighbours after computing the similarities
- For example, KNN is K-documents (or books) in the large number of text documents (books) that are most similar to the queried document

Collaborative Filtering

- Collaborative filtering for frequent itemsets uses weighted bipartite graph matching

Summary

We learnt:

- Social Network Graphs
- Centralities Metrics
- Closeness and sparseness
- Discovering anomaly, ego-net and spam (using near star) from the analysis

Summary

We learnt:

- Page Rank Analysis
- K-cores and
- K-neighbourhoods

End of Lesson 10 on
**Social Network Graph,
Centralities, PageRank and KNN
Analysis**