Evaluation and datasets

Random walks on graphs are often used for evaluation and experimentation in social computing research. To assess the effectiveness of algorithms and models, researchers use specific datasets and evaluation metrics.

What is social media data?

Social media data provides commonly available information on social media channels published by social media users. These data points are in the form of blogs, posts, likes, followers, clicks, shares (reposts and retweets), comments, or engagement rates. Typically, marketers use this data to obtain audience insights, sentiment analysis, and for target audience segmentation. Social media analysis can also provide some indication of the level of brand awareness and customer satisfaction, which can help measure the effectiveness of marketing campaigns and social media strategy.

Evaluation of Random Walks on Graphs:

Evaluation Metrics:

Accuracy: Measures how well the random walk-based algorithm or model predicts user behavior, recommendations, or network properties.

Precision and Recall: Assess the relevance and completeness of recommendations or community detection results.

F1-Score: Combines precision and recall to provide a balanced evaluation metric.

AUC-ROC: Measures the area under the Receiver Operating Characteristic curve, often used for link prediction and recommendation tasks.

Mean Average Precision (MAP): Evaluates the quality of ranked lists in recommendation systems.

Coverage: Indicates how many users or items are covered by the recommendations or community detection results.

Diversity: Measures the variety of recommendations or communities to ensure serendipity and avoid redundancy.

Cross-Validation:

Random walks on graphs are often evaluated using cross-validation techniques to divide data into training and test sets, ensuring that models generalize well to unseen data.

Comparison to Baselines:

Random walk-based methods are compared to baseline models to determine if they offer significant improvements in various social computing tasks.

Datasets in Social Computing:

1. **Online Social Networks:**

Facebook: Several datasets from Facebook have been used in social computing research to study user behavior, interactions, and information diffusion.

Twitter: Twitter datasets are popular for analyzing trends, sentiment analysis, and network dynamics.

LinkedIn: LinkedIn data is used for professional network analysis, recommendation systems, and user profiling.

2. **Collaborative Filtering:**

Movielens: A well-known dataset for movie recommendation research, containing user ratings and movie metadata.

Amazon Product Reviews: Amazon provides datasets of user reviews and product information for various categories.

3. **Information Diffusion:**

Twitter Diffusion Dataset: Contains information about the spread of tweets and hashtags on Twitter, enabling research into information diffusion processes.

4. **Link Prediction:**

Cora: A dataset for link prediction in citation networks, used for academic paper recommendation research.

LinkedIn Connections: LinkedIn datasets with information on user connections, often used for link prediction studies.

5. **Semantic Web and Knowledge Graphs:**

DBpedia: A semantic knowledge graph derived from Wikipedia, used for various information retrieval and semantic computing tasks.

WordNet: A lexical database for the English language, used in natural language processing and text mining.

6. **Web Graphs:**

Web Data Commons: Datasets collected from the web, used for studying web graph properties, link analysis, and information retrieval.

7. **Healthcare and Biomedical Data:**

MIMIC-III: A dataset containing healthcare data from critical care units, used for health informatics research.

PubMed: Provides access to a vast collection of biomedical literature, used for text mining and information retrieval.

These datasets are used to evaluate the performance of random walk-based algorithms and models in various social computing tasks, including recommendation systems, information diffusion modeling, community detection, and more. Researchers often choose datasets that are relevant to their specific research questions and objectives.