Scalability Issues in Online Social Networks

Pedro Neto (pjsn@cin.ufpe.br)

INTRODUCTION

1. A social network is a platform that facilitates individuals to communicate with each other, connected through social relations, such as family, friends, and colleagues [Nepali and Wang 2014; Symeonidis et al. 2014].
2. The enormous and rapid growth of social networks gives rise to novel challenges, particularly related to scalability
3. Scalability refers to the ability of a system to maintain the performance under an increased load.

Social Network Architectures

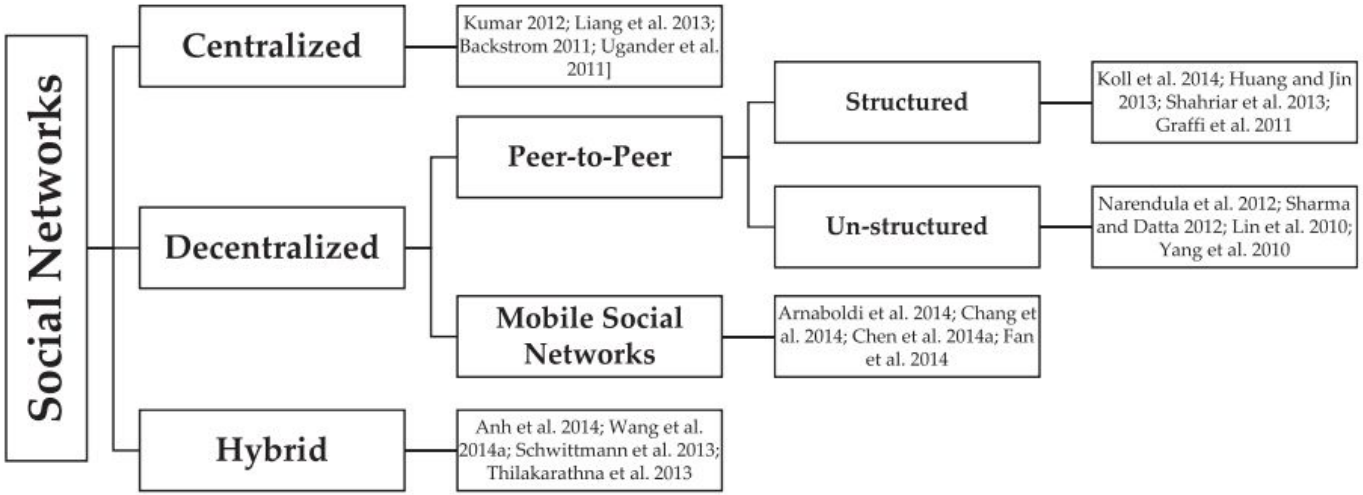

Fig. 1. Taxonomy of social network architectures.

Social Network Architectures

- Centralized Social Network
  - users typically connect to and use various services of social networks through a Web browser
  - based on client–server architecture and are controlled by a single administrative authority
  - Ex.: Facebook, Twitter
- Decentralized Social Network
  - uses the storage space of its participants to increase the availability or the survivability of the data.
  - P2P networks
    - enable the participants to share resources
    - P2P systems have the ability to self-configure, especially in the case of transient failures.
    - Used extensively for file sharing. Ex.: BitTorrent
  - Mobile social networks
    - allow users to communicate and share content with other mobile users in the physical proximity using Bluetooth or Wi-Fi technology without using the Internet
- Hybrid Social Network

SCALABILITY IN SOCIAL NETWORKS

Scalability is very crucial for social networks to provide consistent performance, especially in the case of rapid increase in the number of users

Scalability Issues in Centralized Social Networks

- Large Number of Highly Connected Users
  - Ex.: the tweet and retweet from a celebrity with millions of followers generates more traffic
- Infrastructure Issues
  - require huge infrastructure for thousands of servers
  - Cost of Equipment, Operational Expenses, Energy Consumption
- Two solutions for scaling the infrastructure
  - vertical scaling
    - addition of computational resources, such as processors, memory, and storage, to existing servers or virtual machines or acquiring new servers
  - horizontal scaling
    - In the case of cloud computing, horizontal scaling refers to the process of adding more virtual machines to the existing pool of resources [Yang et al. 2014].
  - Facebook, Twitter, and LinkedIn, have adopted horizontal
- Internal Network Traffic

Scalability Issues in Centralized Social Networks

- User-Generated Content Management and Dissemination
  - Case about retweet of a famous person
- Database Scalability
  - Unstructured data
  - Some of the scalable data storage solutions
    - Cassandra, Haystack, Bigtable, MongoDB

Scalability Issues in Decentralized Social Networks

- Profile and Content Availability
  - ensure availability of users' profiles and UGC.
- Content Distribution
- Energy Efficiency
  - One of the most precious resources for portable devices is energy
  - The excessive use of limited battery presents a key obstacle in the wide adoption of decentralized social applications
- Security and Privacy
  - Providing access to users' data while ensuring security requirements, such as confidentiality, integrity, authentication, access control
- Large-Scale Implementation
  - there is no large-scale implementation of decentralized social networks.
  - The closest was a social network called Diaspora

Scalability Issues in Hybrid Social Networks

In hybrid social network architectures, usually a few central servers are required that can store only necessary information, such as user authentication data, content indexing, and user location.

In most cases, user-generated content is stored on user devices.

Hybrid architectures don’t require large database storage systems to store and retrieve user content.

Scalability Metrics for Social Networks

Table VIII. Scalability Metric of Social Network Architectures

Parameters	Centralized	Decentralized	Hybrid
Availability	✓	✓	✓
Latency	✓	✓	✓
Interserver Communication	✓	✗	✓
Cost of Resources	✓	✗	✓
Cost of Engineering	✓	✗	✗
Server Energy Consumption	✓	✗	✓
Infrastructure Maintenance Cost	✓	✗	✓
Internet Bandwidth Requirement	✓	✗	✓
Data Consistency	✓	✗	✓
Data Replication	✓	✓	✓
Privacy	✓	✓	✓
Security	✓	✓	✓

CONCLUSIONS

Scalability is an important parameter to determine the long-term efficiency and effectiveness of modern large-scale social networks