Small-world phenomena in paper networks

Presenter: 方 曦

2017.5.22

Outline

- Overview
- Introduction
- Work
- Future work

Overview

What is the study?

Small-world

Paper network

Whether small-word or a similar phenomenon occurs in paper reference networks.

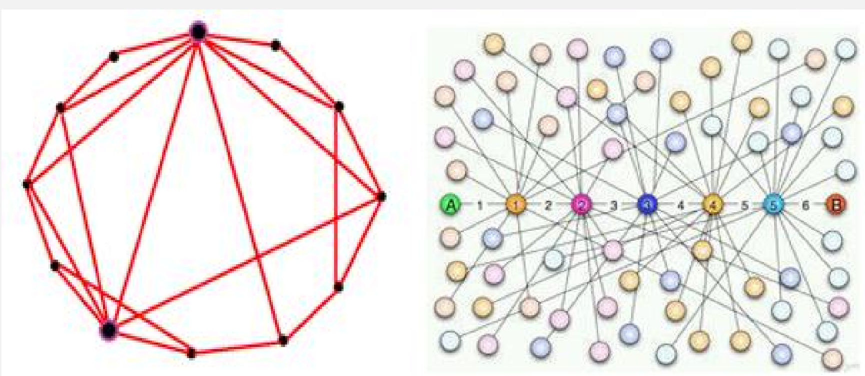
Main point:

Analyze the characteristics of paper network (especially the path in the paper references net)

Introduction

• What is small-word

Small-world network is a type of mathematical graph in which most nodes can be reached from every other node by a small number of hops or steps.



• Small-word theory

Small-world networks and management science research: a review (2007)

history & theory

Two Important Properties of Small World Networks:

① Low average hop count (L)
② High clustering coefficient (CC)

Work

Dataset

Tools

Result

Analysis

Microsoft Academic Graph(MAG) text networks

http://acemap.sjtu.edu.cn/acenap/index.php/datasets.html

Microsoft Academic Graph(MAG) text networks

Name	Nodes	Edges	Communities	VocabularySize
01_MAG_Internet_privacy	749	749	346	85
20_MAG_Database	323411	323411	15696	9000
21_MAG_World_Wide_Web	348927	348927	14580	9268
22_MAG_Computer_network	380102	380102	14619	8498
23_MAG_Parallel_computing	395579	395579	15648	9680
24_MAG_Embedded_system	399925	399925	13617	8908
25_MAG_Algorithm	698494	698494	19721	13900
26_MAG_Operating_system	777301	777301	20227	15469
27_MAG_Programming_language	991394	991394	22995	22453
28_MAG_Telecommunications	1062077	1062077	20347	18214
29_MAG_Artificial_intelligence	1089521	1089521	21984	20260
30_MAG_Computer_vision	1167609	1167609	19635	19142
31_MAG_Machine_learning	1505772	1505772	21332	26261
32_MAG_new_Bioinformatics	1840655	1840655	17805	45188

use

749 nodes
749 edges

Work

Dataset

Tools

Result

Analysis

Python3

Python package:

NetworkX

matplotlib

Seaborn: statistical data visualization

Work

- Dataset
- Tools
- Result
- Analysis

Directed graph

Internet privacy paper network
(749 nodes and 749 directed edges)

result from my program:

1. existing roads is 6108
2. average length of the existing road is 5.925671
3. expected length between two nodes is 3.330856
4. average minimum hop(L) is 2.671415

Work

Dataset

Tools

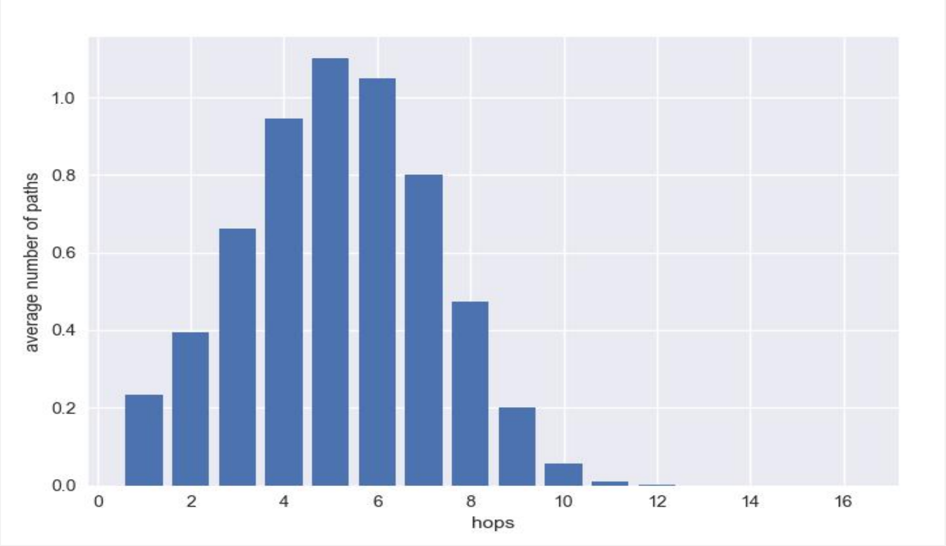
Result

Analysis

Directed graph

Internet privacy paper network
(749 nodes and 749 directed edges)

In the average condition, the hops between arbitrarily two nodes that have paths have a distribution: *visualization the distribution*



Work

Dataset

Tools

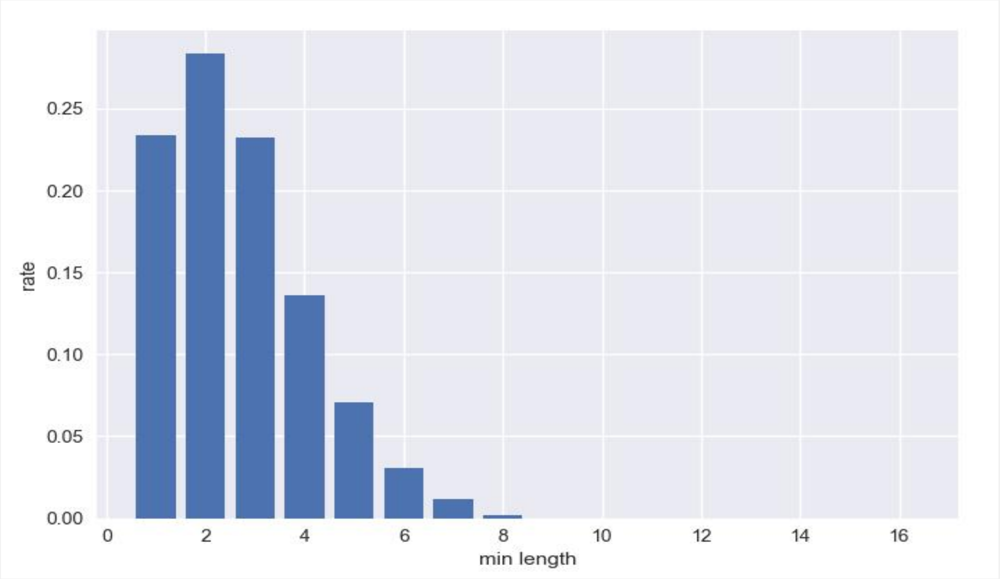
Result

Analysis

Directed graph

Internet privacy paper network
(749 nodes and 749 directed edges)

visualization the distribution of min length



75.0% in 3 hops
98.7% in 6 hops

Work

Dataset

Tools

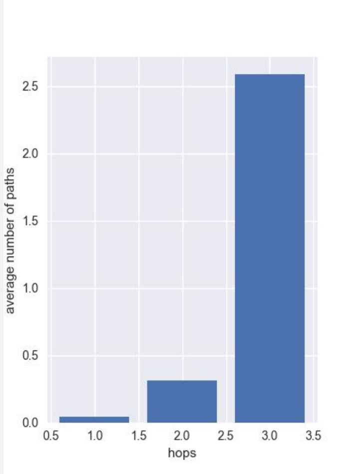
Result

Analysis

Undirected graph

Internet privacy paper network
(749 nodes and 749 directed edges)

In the average condition, the hops between arbitrarily two nodes that have paths have a distribution:(within 3 hops)
visualization the distribution



Work

- Dataset
- Tools
- Result
- **Analysis**

DFS algorithms

the algorithms used now to find all paths between two nodes is a modified depth-first search(DFS).

Algorithm complexity

Find a single path   O(V + E)

but the number of single paths in a graph can be very large, e.g. O(n!) in the complete graph of order n.

Total time the program used may up to O(V!∗ V2)

Future work

- use more powerful computer and improve our algorithms
- calculate clustering coefficient （CC）
- unearth more interesting information from the network

Thanks