Negative Link Prediction in Traffic Recovery

Li xing
2016.11.04

ABSTACT

- We have proposed a way which combines link prediction methods to solve the traffic congestion problem.

- We have found that “weak ties” play a significant role in the efficiency of transportation.

- This work is an exploration of using link prediction method in the traffic problem, it provides a new vision in the traffic control problem and may further affect the link prediction usage in the traffic field.

INTRODUCTION

1.Link prediction(methods, application)
2.Transportation Theory(public transportation, navigation， law)
3.Most of the theory in avoiding traffic congestion are based on the design of traffic routes, but even the best design cannot avoid it completely.So once problem occur， we may relieve it by Link Prediction method.

DATASETS

	|V|	|E|	D	C	<k>	<d>	<H>
BA256	256	508	0.016	0.075	3.97	3.49	2.17
BA512	512	1020	0.008	0.052	3.98	3.68	2.87
BA1024	1024	2044	0.004	0.030	3.99	4.07	2.78
USAir	332	2126	0.039	0.625	12.81	2.74	3.46
Cal Road 10	121	156	0.021	0.142	2.58	9.54	1.13
Cal Road 20	409	525	0.006	0.101	2.57	17.71	1.17
Cal Road 30	1285	1750	0.002	0.077	2.72	23.28	1.16

Origin Data:Road network of California1,965,206 nodes 2,766,607 edges

Negative Recovery Sequence in Link Prediction

1.Nodes with more edges means more important in LP
2.Nodes with more edges requires better processing capability
3. So we use both ways to see what will happen

Negative Recovery Sequence in Link Prediction

1.Preferential Attachment: $s_{xy}^{PA} = k_x \times k_y$ $\quad s_{xy}^{RPA} = \frac{1}{k_x \times k_y}$ $\quad s_{xy}^{RAPA} = \frac{1}{k_x + k_y}$

2.Local Path: $S^{LP} = A^2 + \partial A^3$ $\quad S^{RLP} = \frac{1}{A^2 + \partial A^3}$

3.Similarity Attachment: $S_{xy}^{SA} = \tau(\tau(\tau(x))) \cup \tau(\tau(\tau(y)))$ $\quad S_{xy}^{SA} = \frac{1}{\tau(\tau(\tau(x))) \cup \tau(\tau(\tau(y)))}$

EXPERIMENT

method

1. Create the unweighted & undirected networks
2. Randomly break the edges in the network (**circle?**)
3. Use Link Prediction correlated methods to rank the broken edge
4. Recover the routes in the sequence of the ranked score
5. Conduct package simulation experiments

NOTICE: We think that in the status of congestion the traffic in the route almost Stagnated, so to some degree it equals the route has been broken.

EXPERIMENT

package simulate

1. In each time step, each node generates R/N packets, where N is the size of the network and R is a parameter tuning the generation rate
2. The packets are initialized with random destinations.
3. Use shortest path algorithm to determine the traffic route
4. At each time step, transfer the node i to the next node according to the routing table. When the node cannot transfer all the packets accumulated in its queue, it deals with them following the first in first out rule.
5. Calculate the order parameter in the traffic net.

Notice：

1. When R/N is not an integer, we create IntR/N packets determinately and create a packet simultaneously with probability p=R/N−IntR/N.
2. We set a transmission capacity C_i to the node i, i=1,2, ... ,N, which means that, at each time step, the maximal number of packets transferred by the node i to the next node according to the routing table is C_i .Here in this paper, we set $C_i = 1$.

EXPERIMENT

order parameter

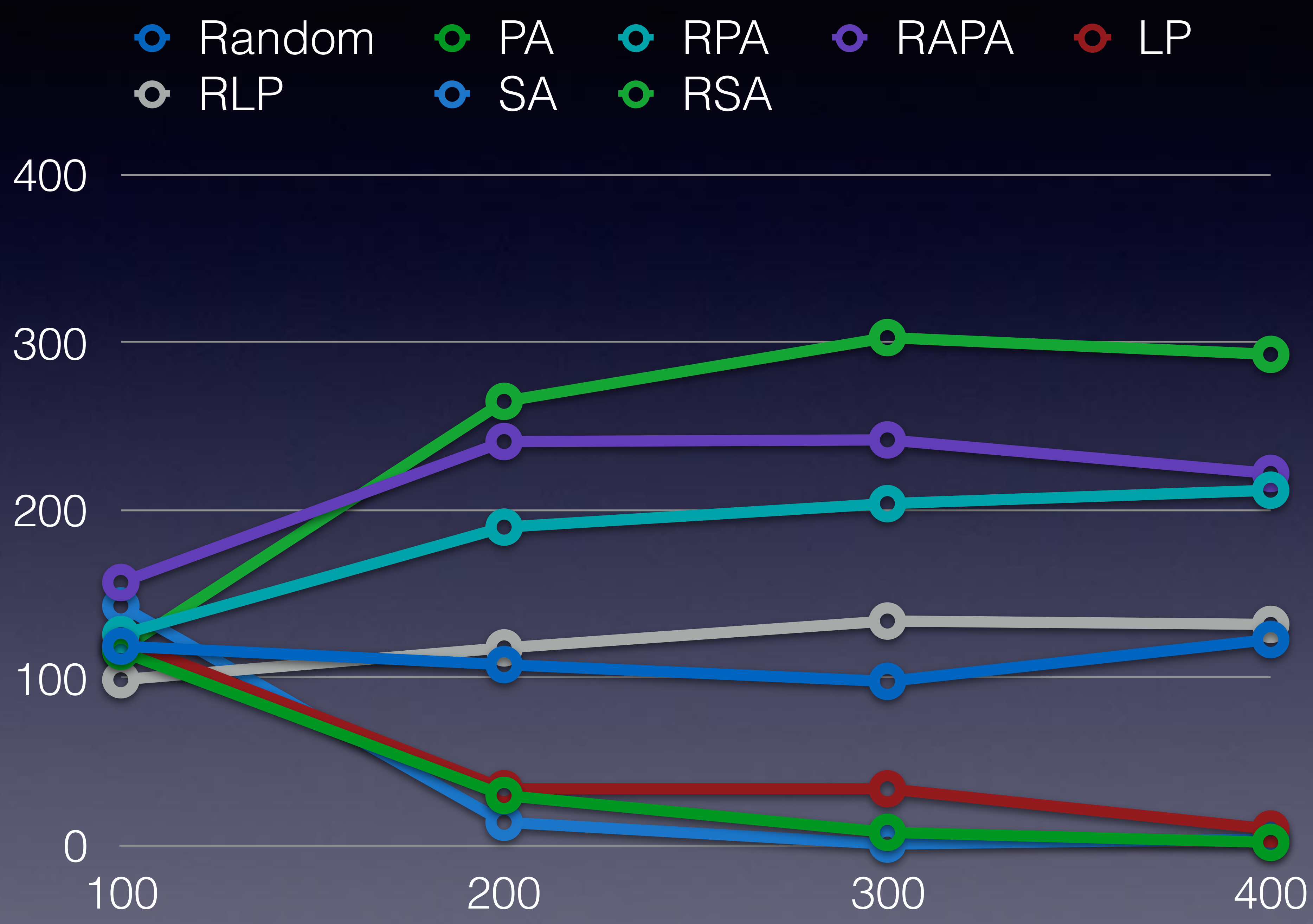
W(t) is defined as the number of packets on the network at time step t

$\triangle W = W(t + \triangle t) - W(t)$

$$u(R) = \lim \frac{< \Delta w >}{R \Delta t}$$

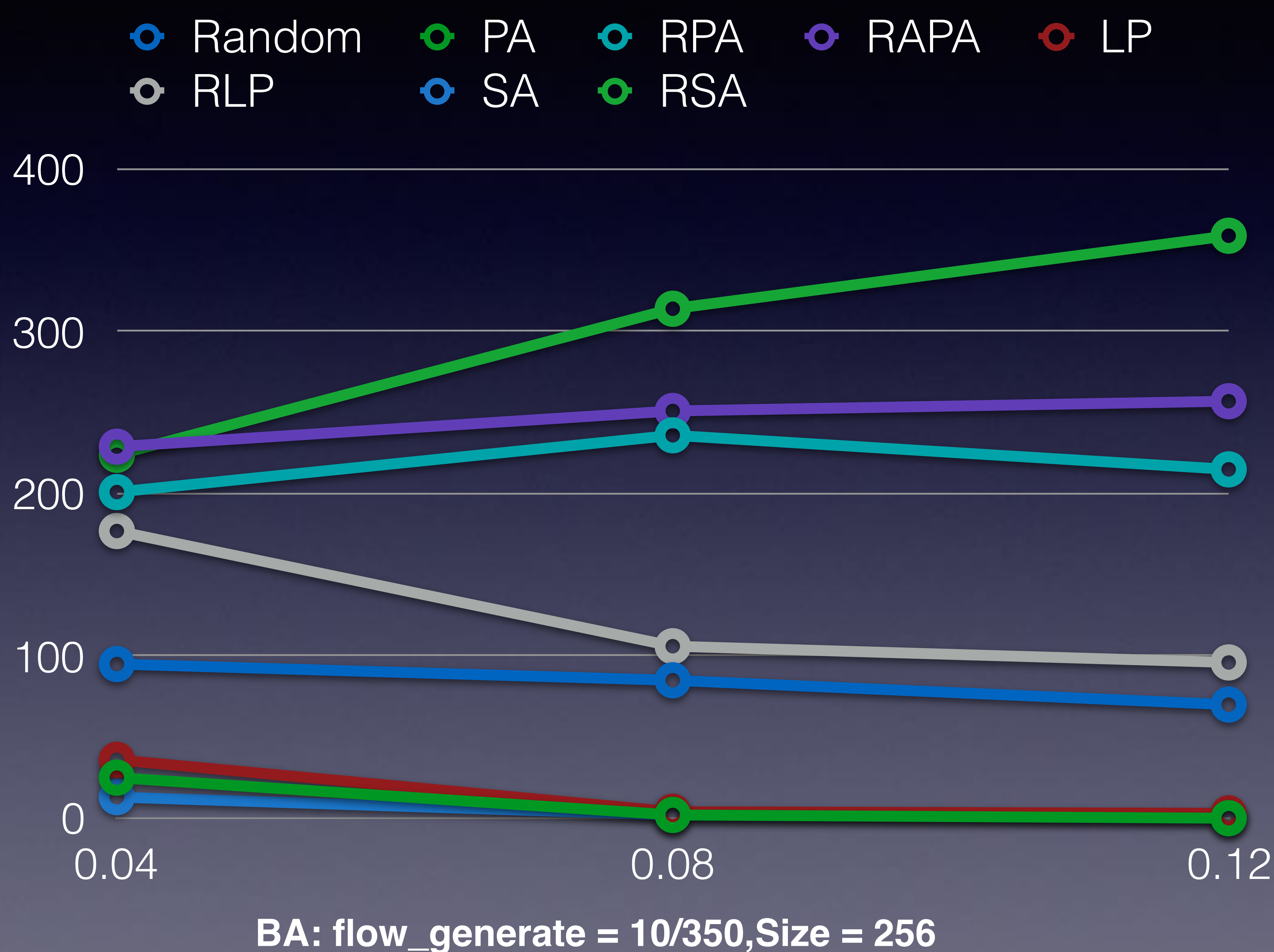
the order parameter represents the ratio between the existing flow and the inflow of packets

EXPERIMENT



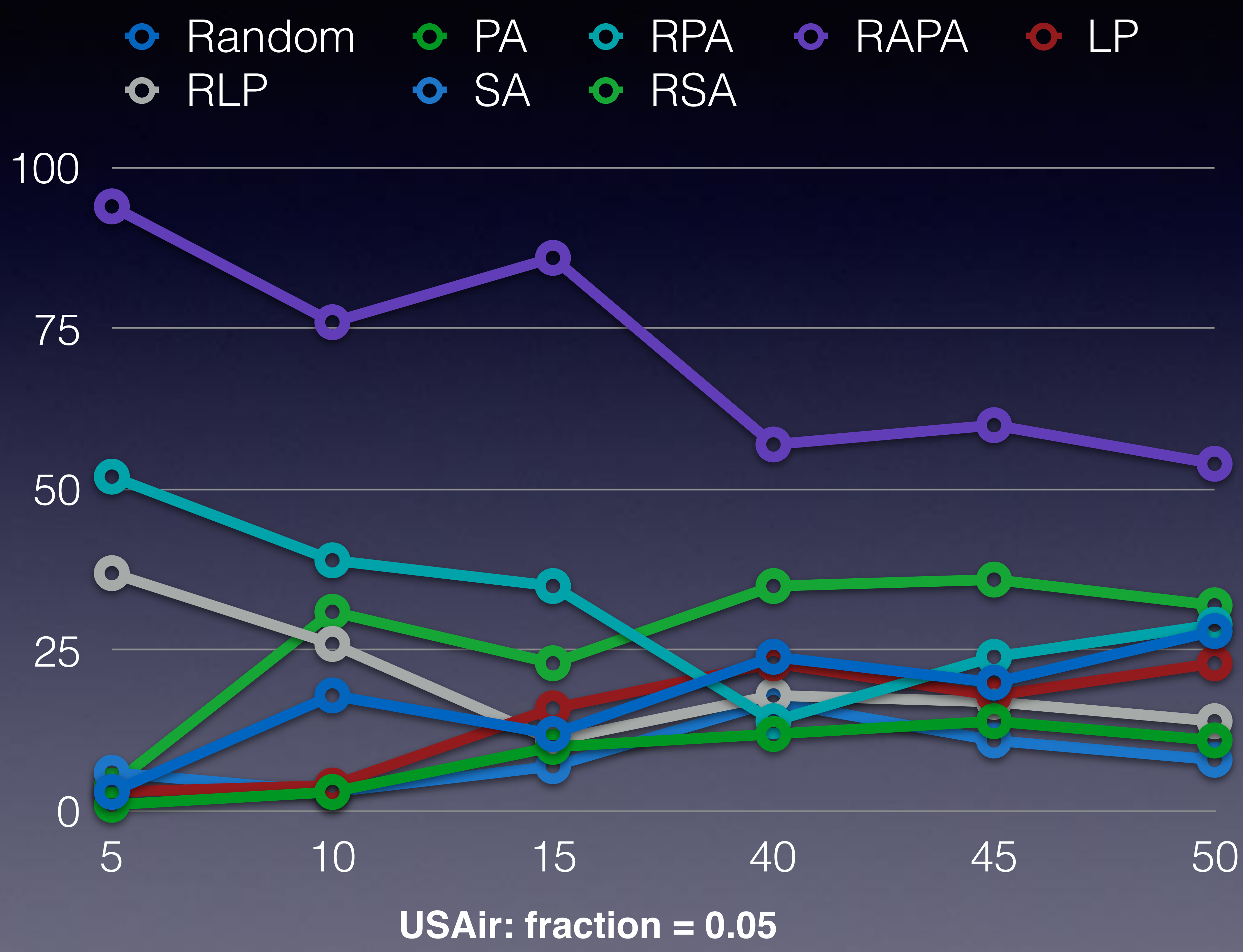
BA: flow_generate = 10/350, fraction = 0.05

EXPERIMENT



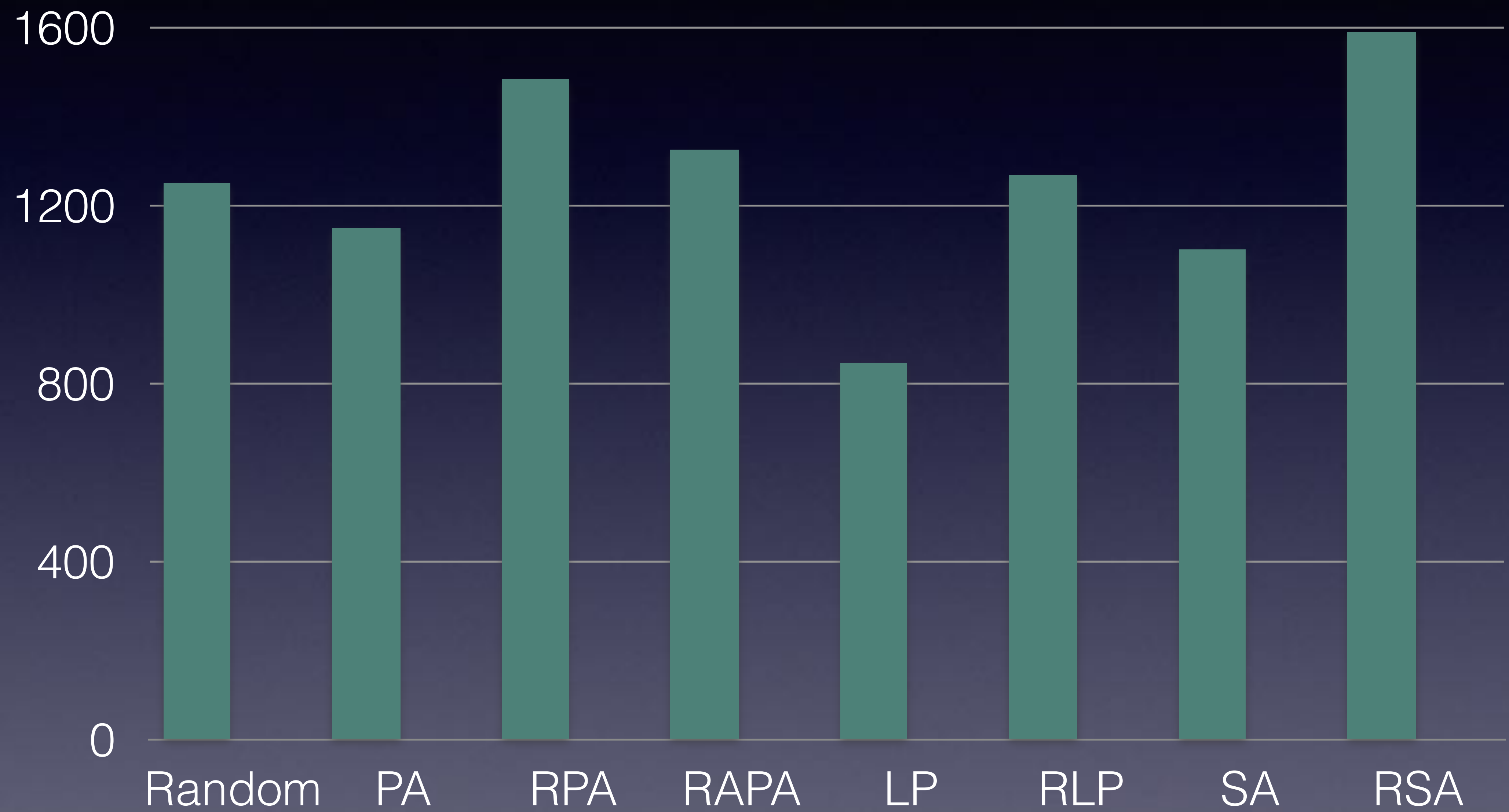
Random
PA
RPA
RAPA
LP
RLP
SA
RSA
400
300
200
100
0
0.04
0.08
0.12
BA: flow_generate = 10/350,Size = 256

EXPERIMENT



Random
PA
RPA
RAPA
LP
RLP
SA
RSA
100
75
50
25
0
5
10
15
40
45
50
USAir: fraction = 0.05

EXPERIMENT



Data:Cal Road 20, Loop = 10000,Fraction = 0.1, flow_generate = 10 / 350.0

CONCLUTION

1.高的度异质性可能导致较好的效果
2.果然是负相关的(弱连接)

Future work:
1.Take route length into account
2.Take node's differ into account
3.New link prediction method in road network

Optimized RA

异常边
度的平均值
pa + cn调参
大型网络圈定范围
弱连接
ba网络，ws，grid，int网络提高准确率
共同邻居信息

Optimized RA

1.CN(common neighbor) $s_{xy} = |\tau(x) \bigcap \tau(y)|$

2.AA(Adamic-Adar Index) $s_{xy} = \sum_{z \subset \tau(x) \bigcap \tau(y)} \frac{1}{\log k_z}$

3.RA(Resource Allocation) $s_{xy} = \sum_{z \subset \tau(x) \bigcap \tau(y)} \frac{1}{k_z}$

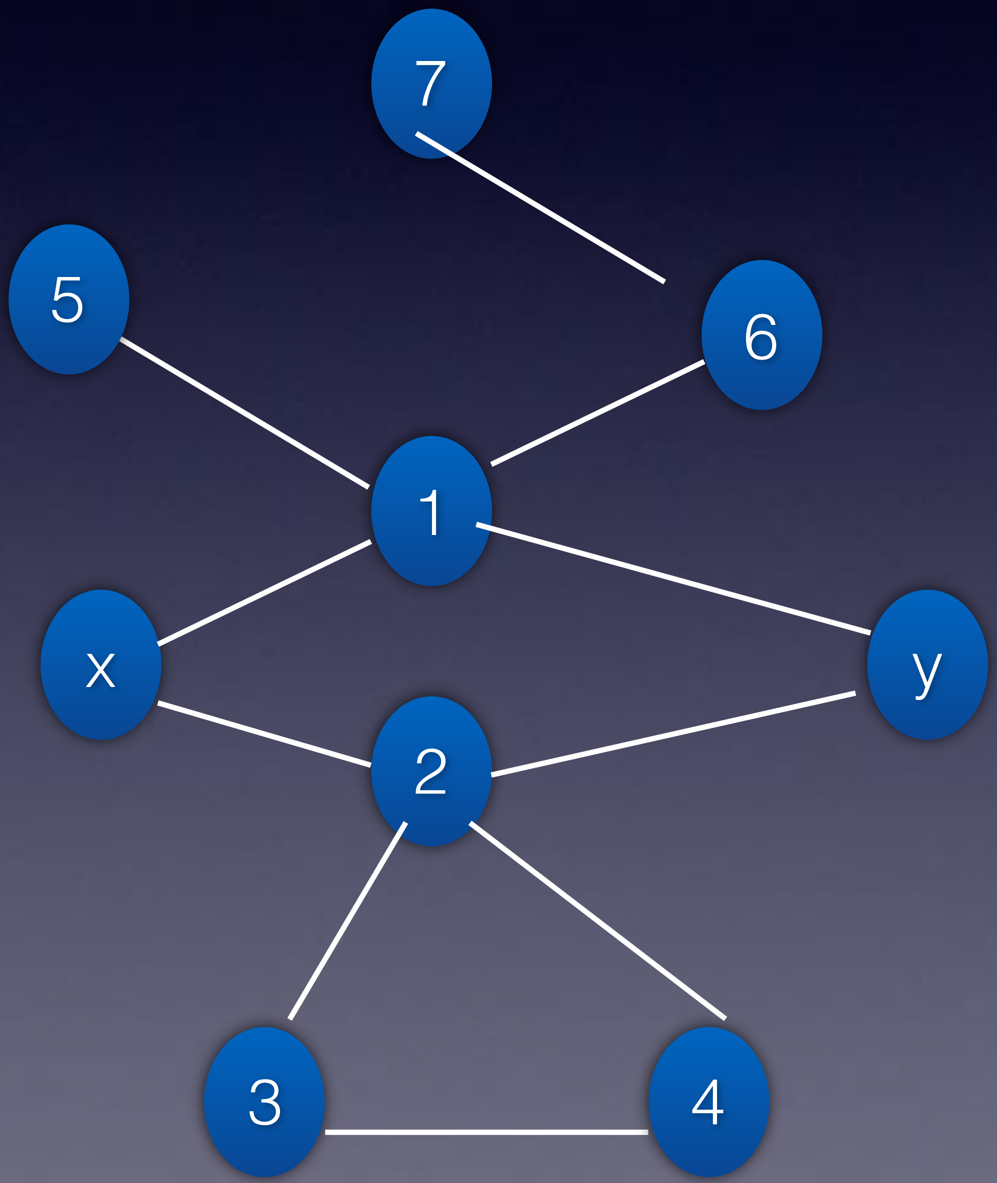
4.ORAN(Optimized Resource Allocation By Neighbor)

5.ORAC((Optimized Resource Allocation By Clustering) $s_{xy} = \sum_{z \subset \tau(x) \bigcap \tau(y)} \frac{c(z)}{k_z}$

6.PA

7.LP

Model



DATASETS

	|V|	|E|	D	C	<k>	<d>	<H>	效果
JAZZ	198	2742	0.1406	0.6175	27.70	2.235	1.40	好
BA	512	1020	0.008	0.048	3.99	3.82	2.39	不好
USAir	332	2126	0.039	0.625	12.81	2.74	3.46	不好
ROUTER	5022	6258	0.0005	0.012	2.49	6.45	5.50	不好
YEAST	2375	11692	0.0041	0.306	9.85	5.10	3.48	不好
FOOTBALL	35	118	0.198	0.339	6.74	2.12	1.49	好
GRID	4941	6594	0.00054	0.080	2.67	18.99	1.45	不好
ELEGANS	297	2148	0.0489	0.2924	14.4646	2.455	1.801	好
SANDI	674	613	0.00270	0.0	1.819	can't reach	1.985	random

DATASETS

	CN	AA	RA	ORAN	ORAC	PA	LP
Elegans	83.65	85.376	85.699	**85.733**	**85.8305**	74.6875	83.6005
Jazz	91.785	92.545	93.635	**93.665**	**93.9**	75.515	91.685
FOOTBALL	64.855	65.2935	**64.8345**	66.066	**66.8895**	70.9995	67.372
SANDI	0.497545	0.497335	0.497375	**0.49758**	**0.5000**	0.36149	0.6201

Round = 1000, Loop = 100
conclusion : ORAC is better

THANKS FOR YOUR TIME !

Q & A