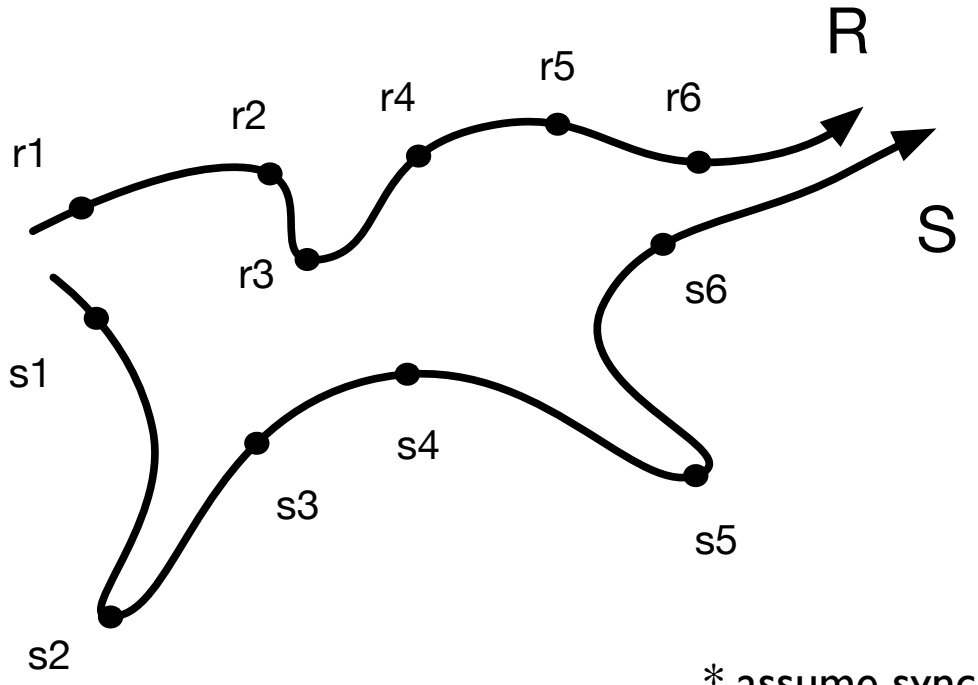
Attraction and Avoidance Detection from Movements

Zhenhui Jessie Li
(with Bolin Ding, Fei Wu, Tobias Lei, Roland Kays, Meg Crofoot)
Pennsylvania State University

VLDB Conference
Hangzhou, China
September, 2014

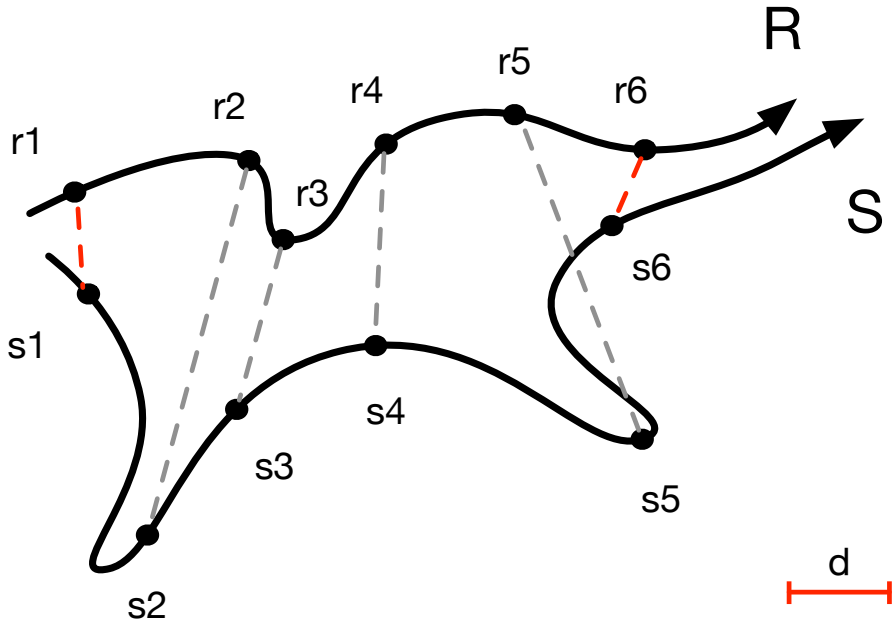
Mining Mobility Relationship Problem

- Given two trajectories R and S, measure their relationship strength



* assume synchronized sampling rate

Using Trajectory Similarity as a Measure of Mobility Strength



Meeting (or co-locating) frequency

$$freq(R,S) = \sum_{i=1}^{n} \tau(r_i, s_i).$$

$$\tau(r_i, s_j) = \begin{cases} 1, & |r_i - s_j| \leq d; \\ 0, & \text{otherwise.} \end{cases}$$

Vlachos et al., Discovering similar multidimensional trajectories. ICDE'02
Chen et al., Robust and fast similarity search for moving object trajectories. SIGMOD'05
Jeung et al., Discovery of convoys in trajectory database. VLDB'08
Li et al., Mining relaxed temporal moving object clusters. VLDB'10

Meeting Frequency = Relationship Strength?

the more frequently you co-locate with another person,

less frequently

the stronger the mobility relationship is.

weaker

Meeting Frequency = Relationship Strength?

Example 1.
A and B are friends
living in different cities
attracted to meet

Freq(A, B) = 2

Example 2.
A and C are colleagues
working in the same building
avoid meeting

Freq(A, C) = 20

Meeting Frequency ≠ Relationship Strength

Consider Mobility Background to Infer Relationship

Example 1.
A and B are friends
living in different cities
attracted to meet

Freq(A, B) = 2

Example 2.
A and C are colleagues
working in the same building
avoid meeting

Freq(A, C) = 20

Mobility background

Expect(A, B) = 1

Expect(A, C) = 100

What happened vs. What is expected to happen

Example 1.	Example 2.	
Freq(A, B) = 2	Freq(A, C) = 20	What happened?
Expect(A, B) = 1	Expect(A, C) = 100	What is expected?

What happened vs. What is expected to happen

Example 1.	Example 2.	
Freq(A, B) = 2	Freq(A, C) = 20	What happened?
larger than	smaller than	
Expect(A, B) = 1	Expect(A, C) = 100	What is expected?
↓	↓	
Attraction	Avoidance	

How to estimate what is expected?

How to Estimate Expectation?

- Null hypothesis: Two movement sequences R and S are independent.
- If we randomly shuffle the sequences,

$$R \rightarrow \sigma(R) \quad S \rightarrow \sigma(S)$$

- the meeting frequency should remain the same

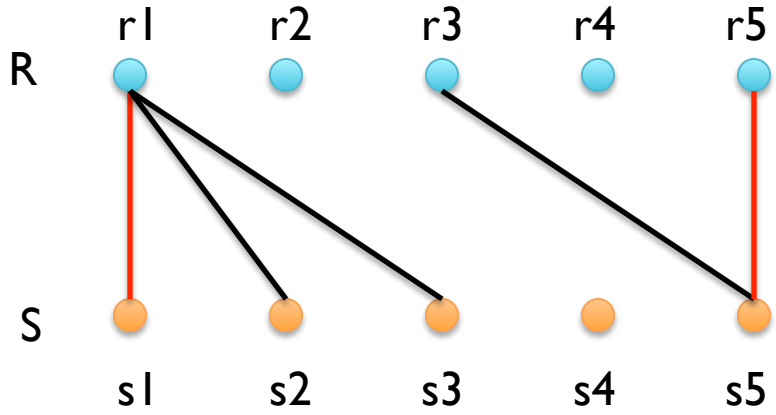
$$freq(R, S) \approx freq(\sigma(R), \sigma(S))$$

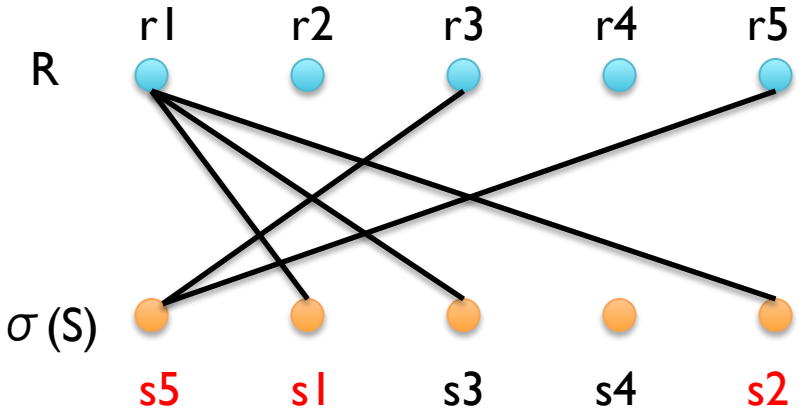
$$Pr(freq(\sigma(R), \sigma(S) = y)) = Pr(freq(R, \sigma(S) = y))$$

Shuffling two sequences = Shuffling one sequence

Permutation Test to Estimate the Probabilistic Background Model

If we randomly shuffle the sequence …





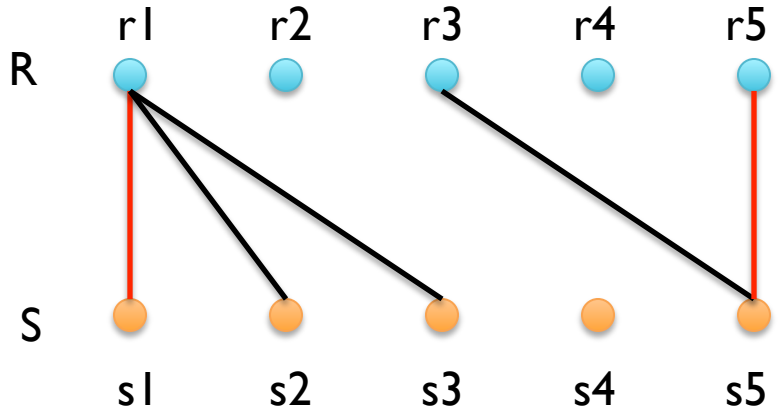
$$freq(R, S) = 2$$

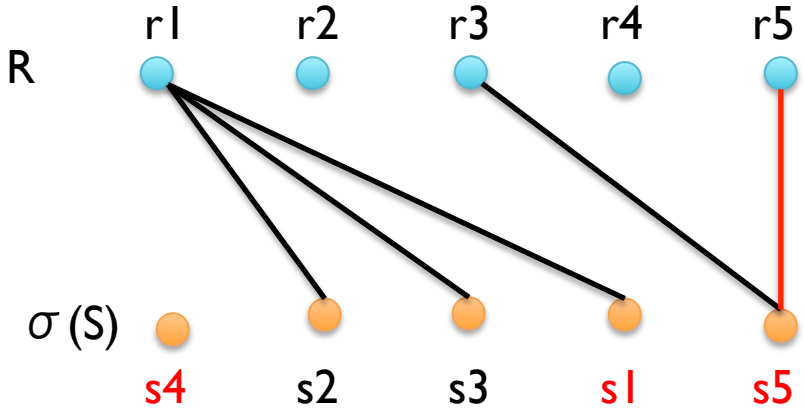
$$freq(R, \sigma(S)) = 0$$

Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Permutation Test to Estimate the Probabilistic Background Model

If we randomly shuffle the sequence …

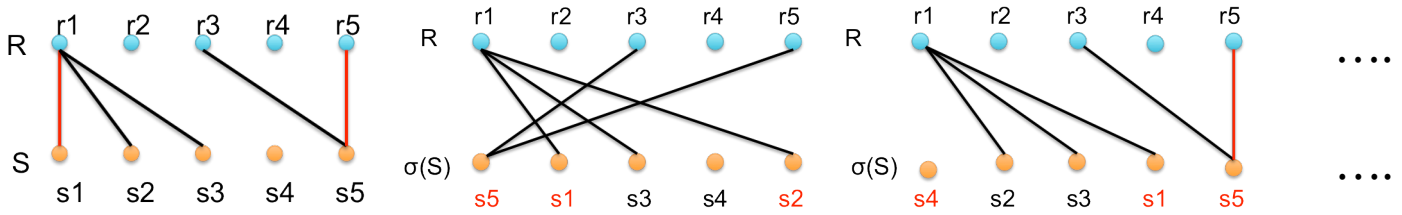




$$freq(R, S) = 2$$

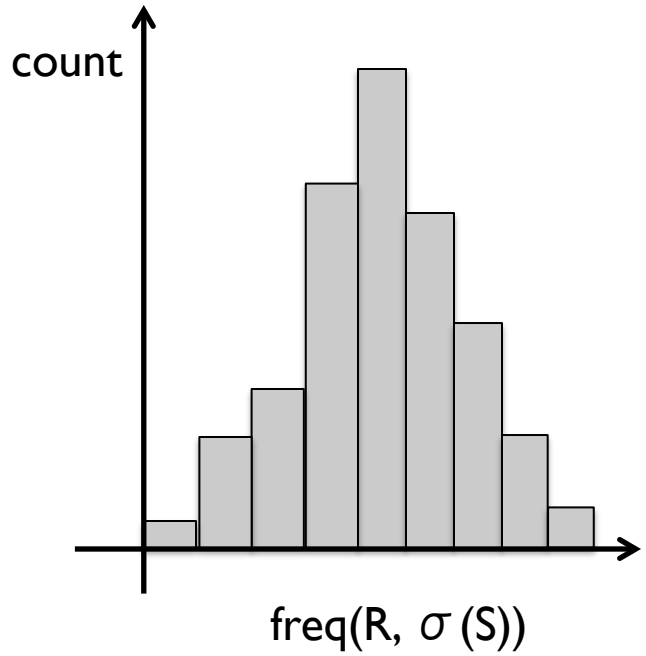
$$freq(R, \sigma(S)) = 1$$

Permutation Test to Estimate the Probabilistic Background Model



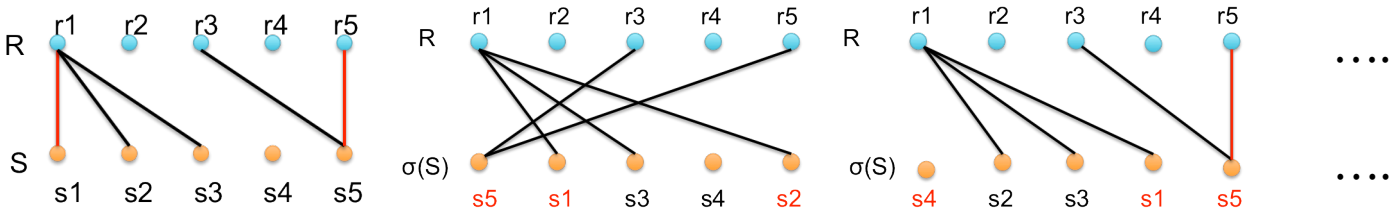
n! permutations

generate histogram



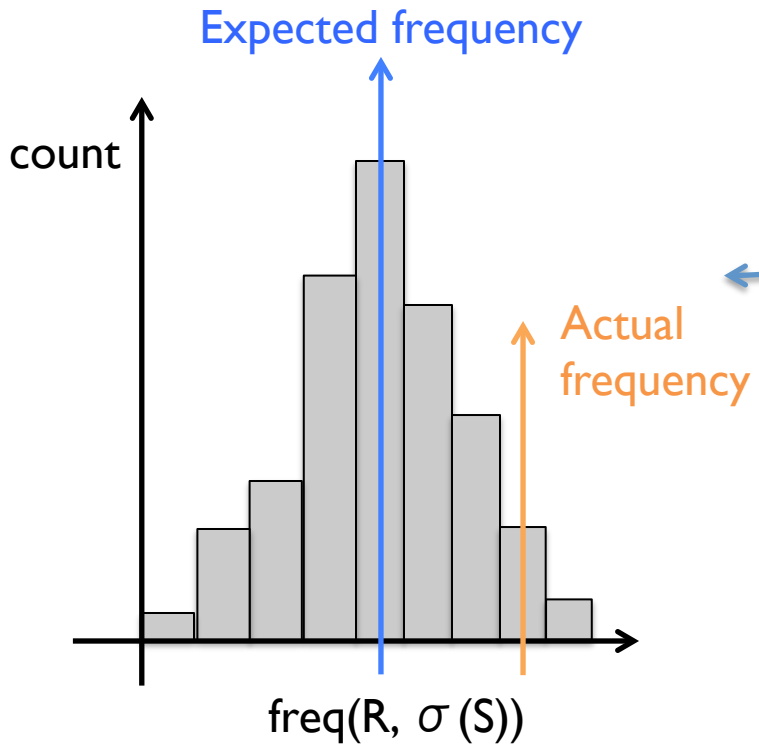
Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Compute Degree of the Relationship



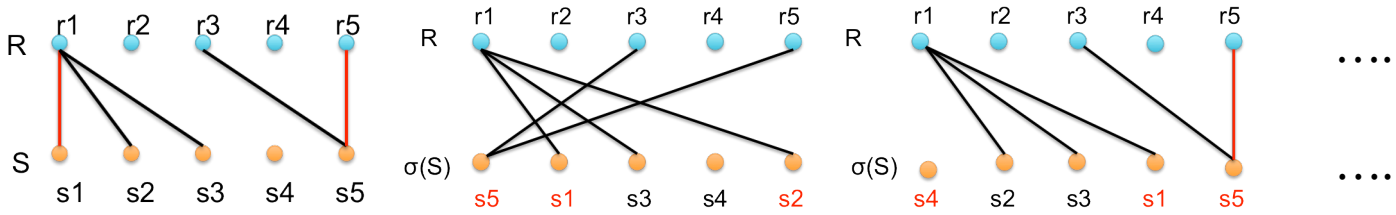
n! permutations

generate histogram



Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

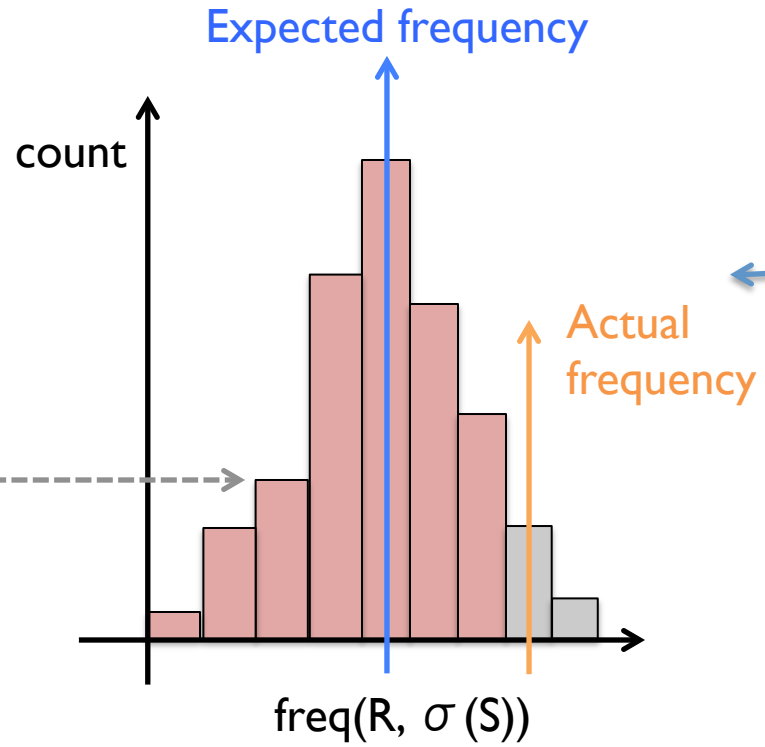
Compute Degree of the Relationship



n! permutations

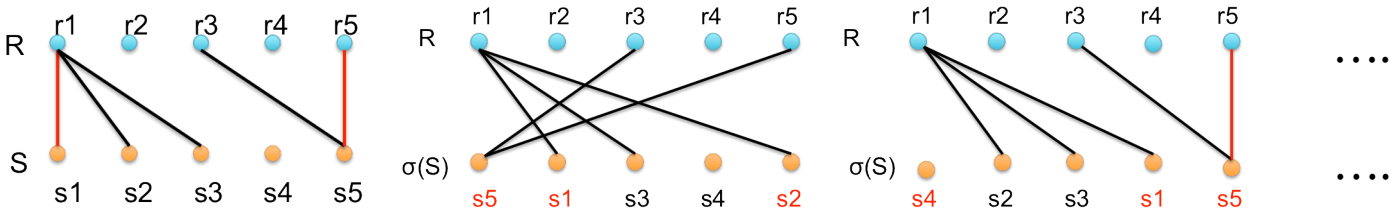
generate histogram

Attraction relationship:
95% significance



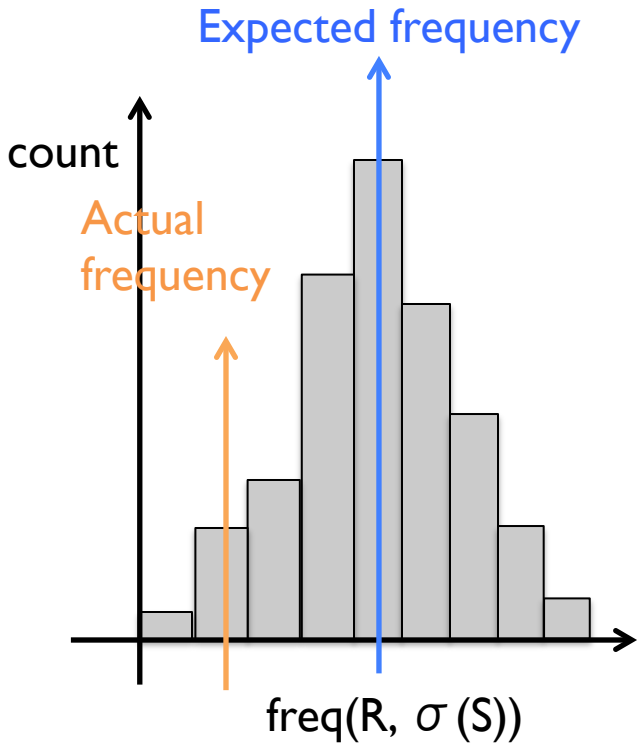
Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Compute Degree of the Relationship



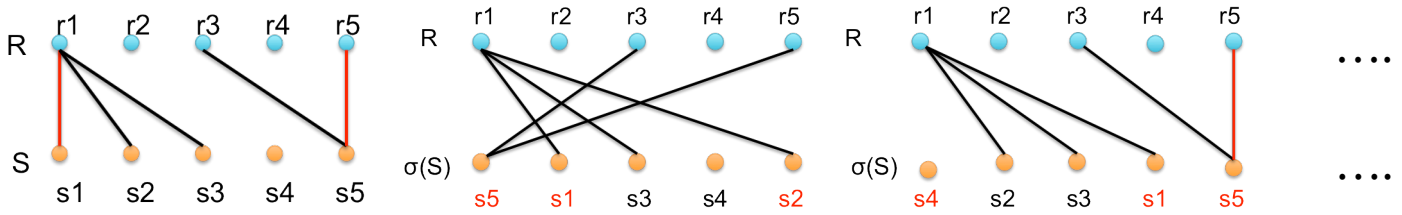
n! permutations

generate histogram



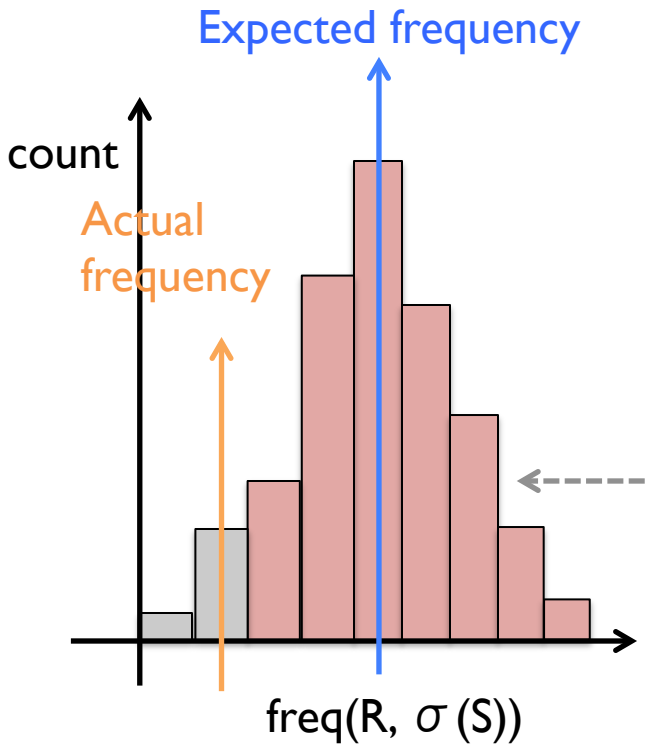
Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Compute Degree of the Relationship



n! permutations

Avoidance relationship:
98% significance

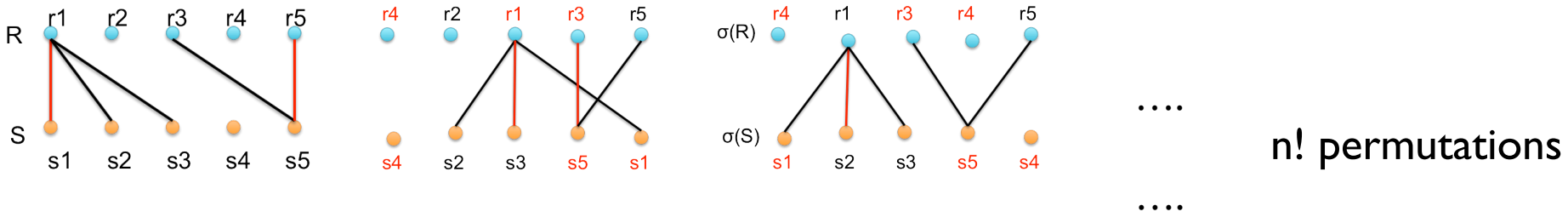


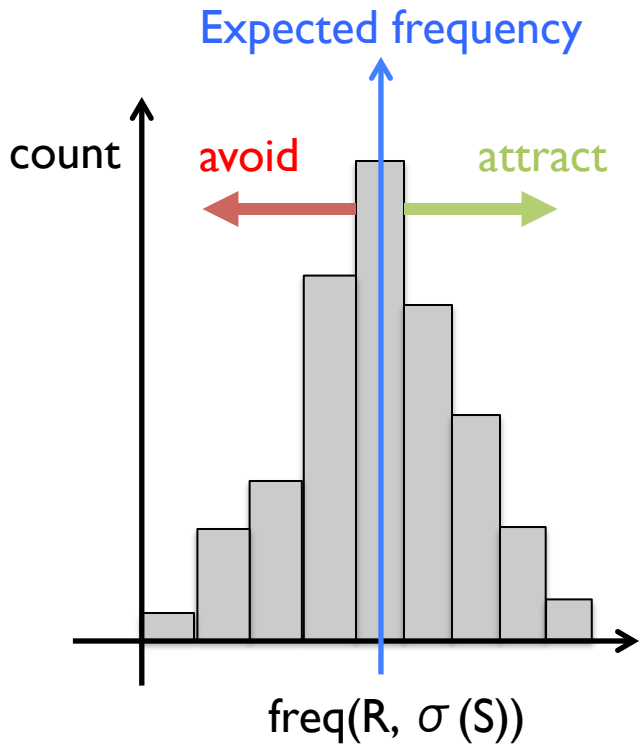
generate histogram

98% area

Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

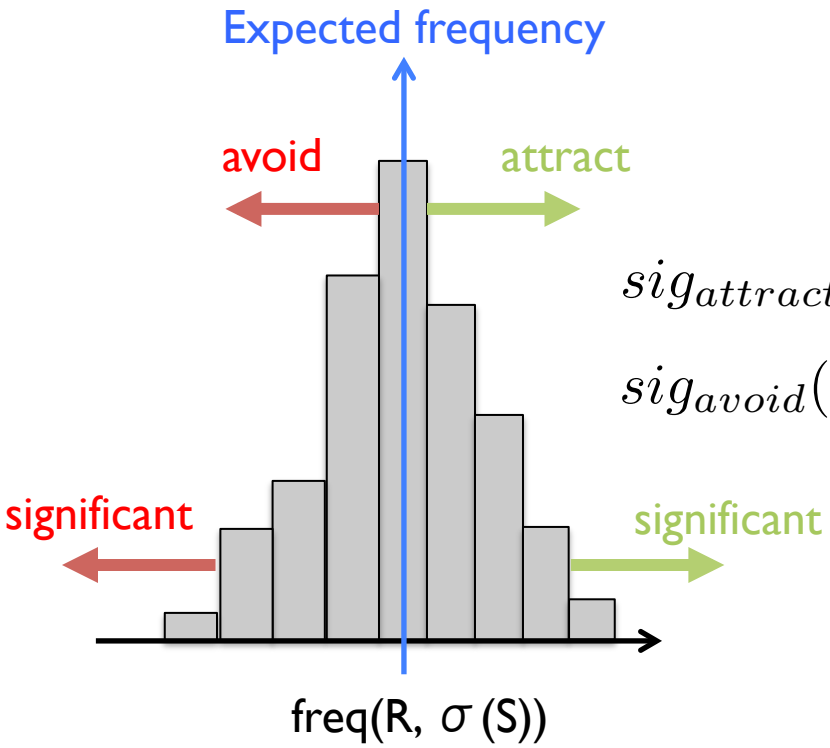
Compute Degree of the Relationship





Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Compute Degree of the Relationship



$$sig_{attract}(R,S) = Pr[freq(R,S) > freq(R,\sigma(S))]$$

$$sig_{avoid}(R,S) = Pr[freq(R,S) < freq(R,\sigma(S))]$$

Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Monte Carlo Scheme to Approximate Degree

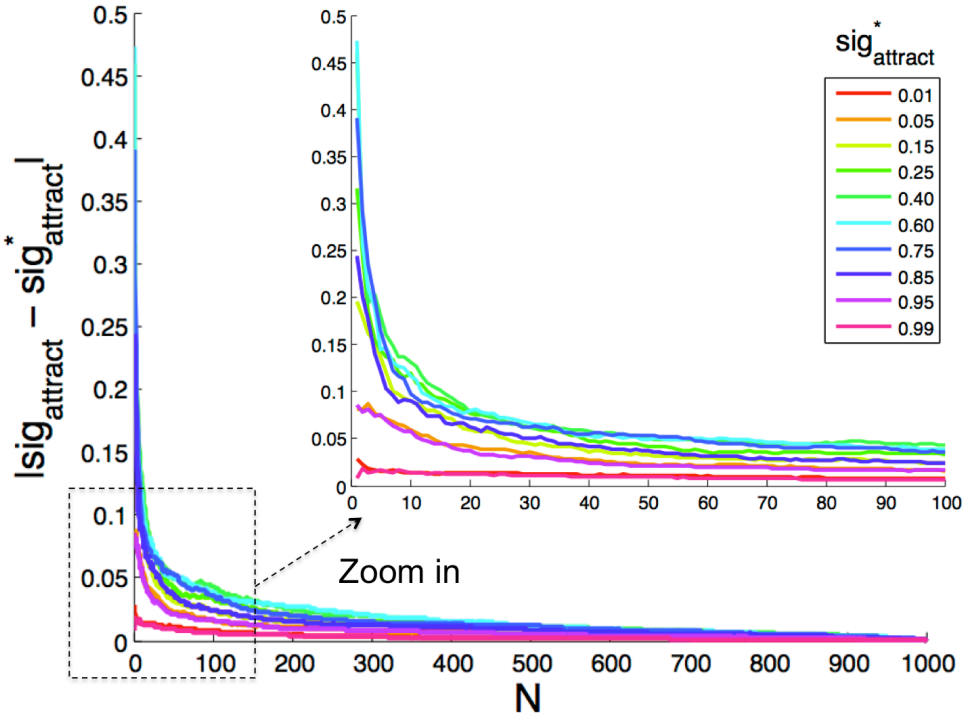
- The total number of permutations is factorial n!
- Monte Carlo scheme: sample N permutations

$$N \geq \frac{4}{\epsilon^2 \rho} \ln \frac{2}{\delta}$$

guarantee

$$(1-\epsilon)\rho \leq \hat{\rho} \leq (1+\epsilon)\rho$$

$$\text{with probability } 1-\delta$$

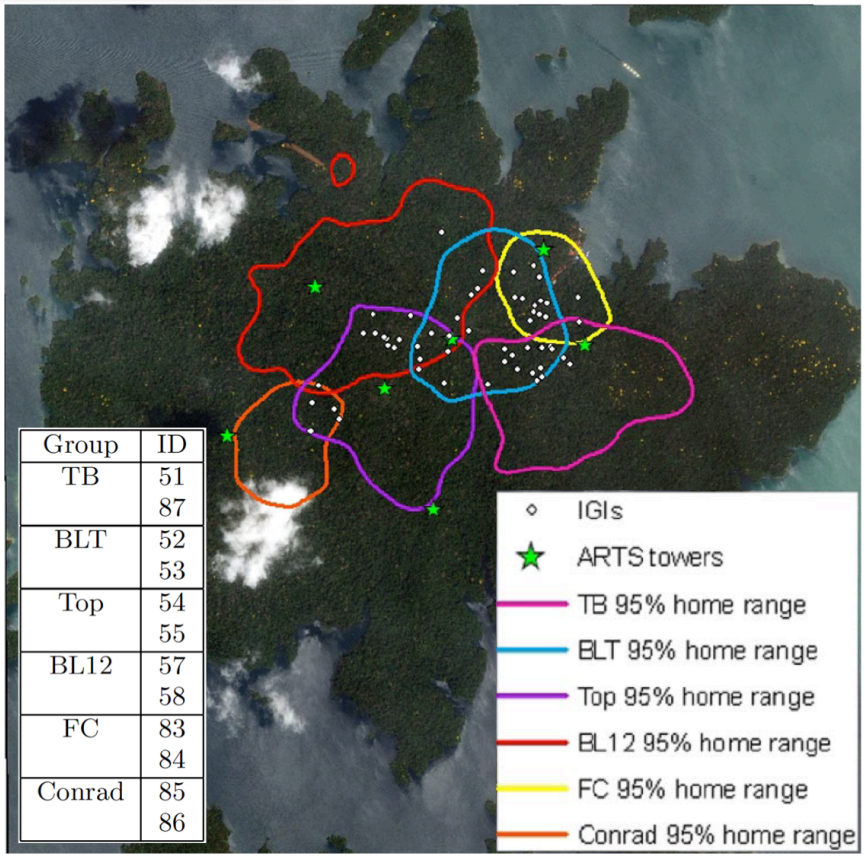


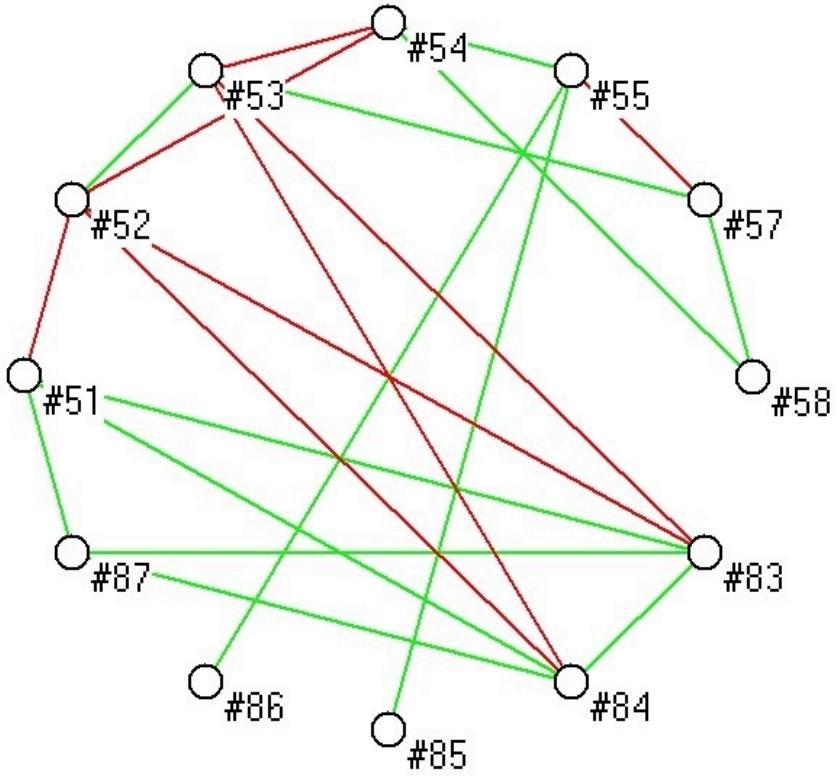
Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Experiment on the Monkey dataset

12 monkeys
11/10/2004 – 04/18/2005

* green line: sig_{attract} > 0.95
* red line: sig_{avoid} > 0.95





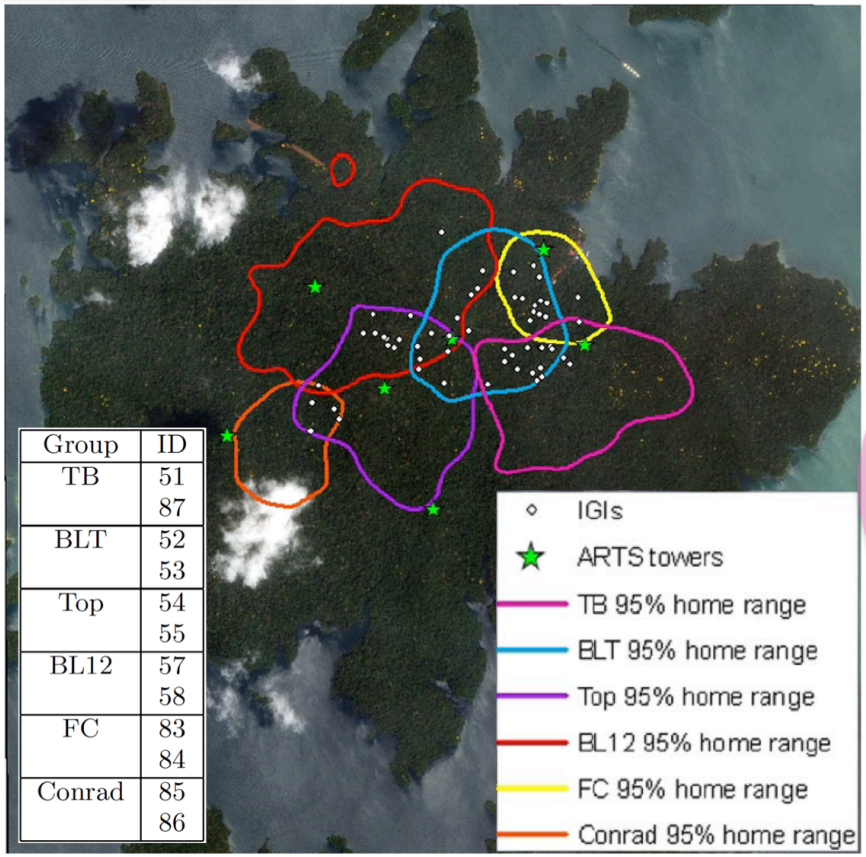
Red: significant avoidance
Green: significant attraction

Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

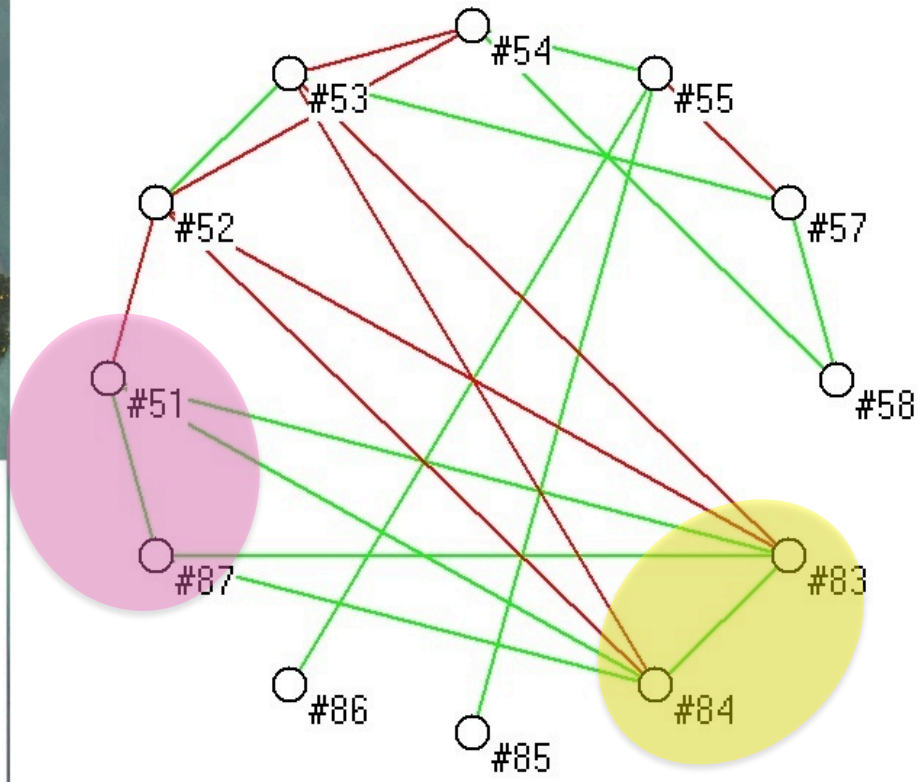
Experiment on the Monkey dataset

12 monkeys
11/10/2004 – 04/18/2005

* green line: sig_{attract} > 0.95
* red line: sig_{avoid} > 0.95



Group	ID
TB	51
	87
BLT	52
	53
Top	54
	55
BL12	57
	58
FC	83
	84
Conrad	85
	86



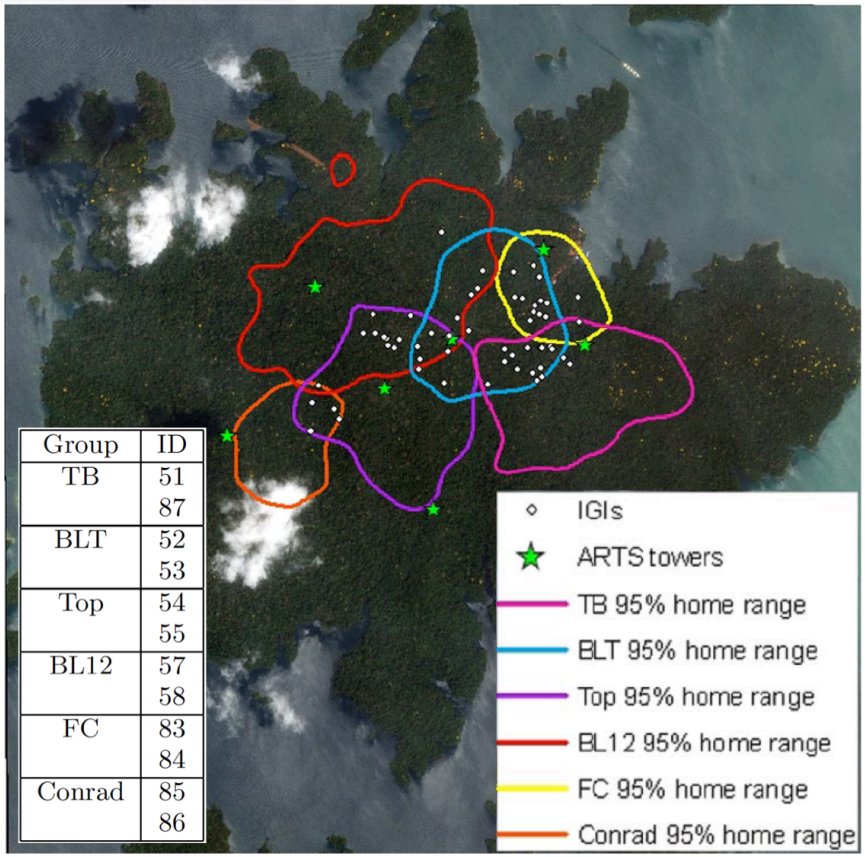
Red: significant avoidance
Green: significant attraction

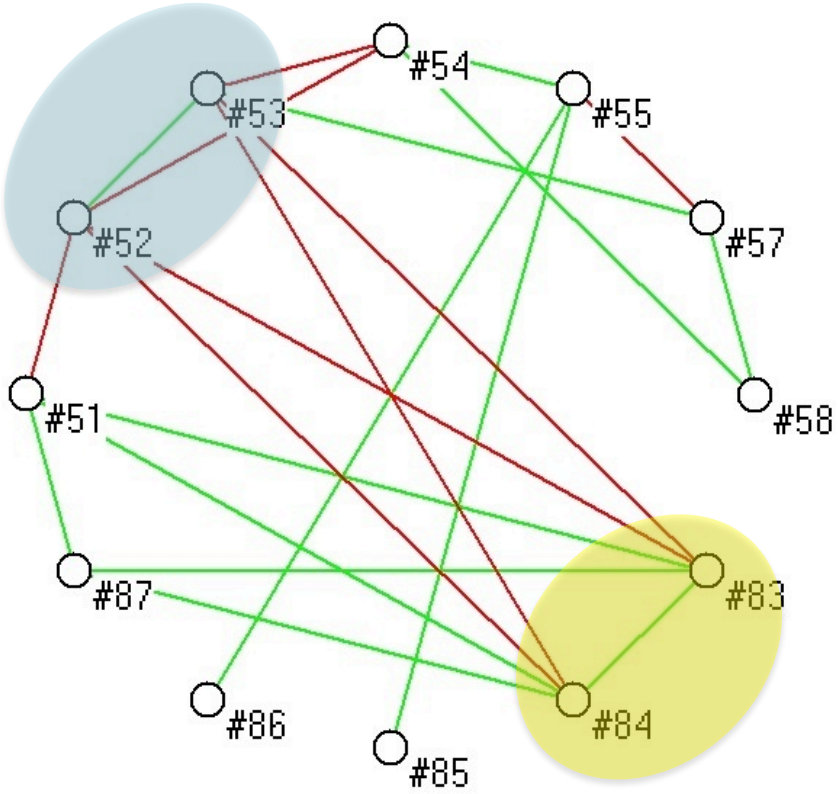
Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Experiment on the Monkey dataset

12 monkeys
11/10/2004 – 04/18/2005

* green line: sig_{attract} > 0.95
* red line: sig_{avoid} > 0.95

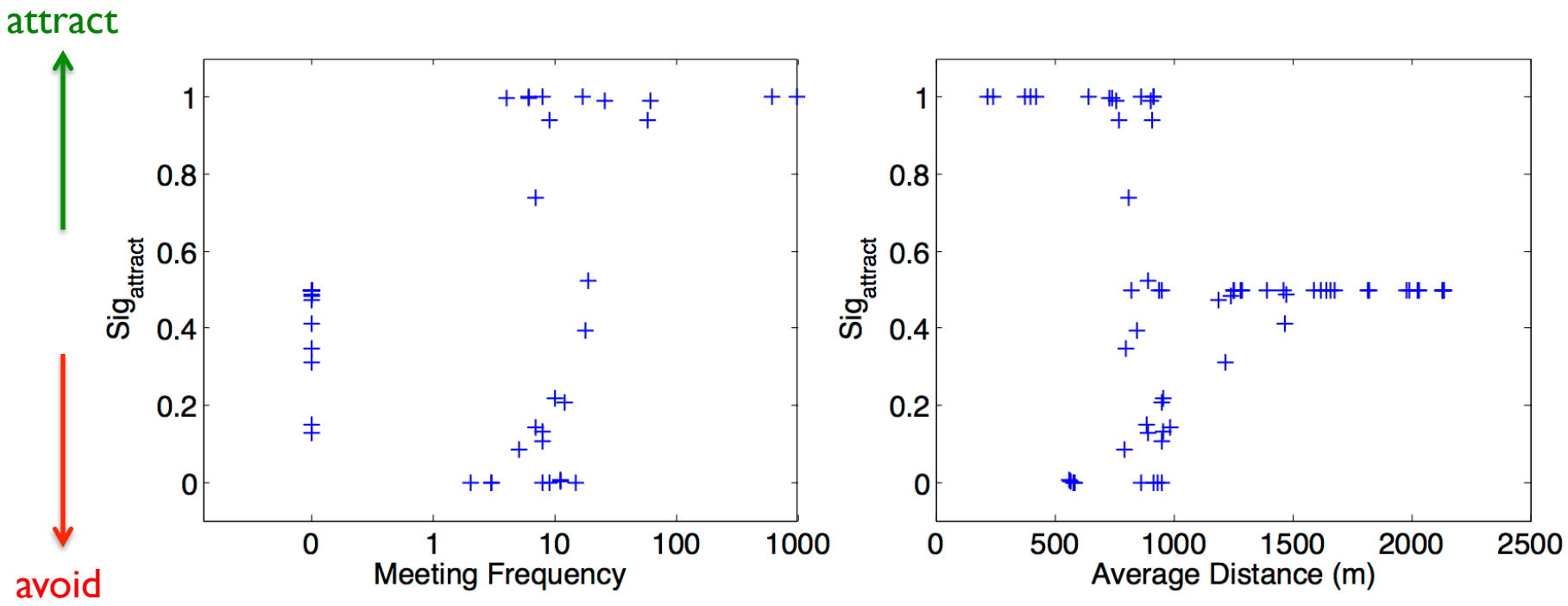




Red: significant avoidance
Green: significant attraction

Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Comparison with Previous Measures



Z. Li et. al., Int. Conf. on Very Large Data Bases (VLDB'14/PVLDB)

Summary and Future Work

- Summary: Important to consider *background*
  - What happened vs. What is expected to happen
  - Consider mobility background using permutation test

Thanks! Questions?

  - How to deal with "impossible" trajectory?
  - How to deal with sparse observations?
- Rich spatial and temporal context
  - location semantics
  - social events