

Finding Self-similarity in Opportunistic People Networks

Ling-Jyh Chen¹, Yung-Chih Chen¹,
Tony Sun²

Paruvelli Sreedevi¹, Kuan-Ta Chen¹
Chen-Hung Yu³, Hao-Hua Chu³

¹Academia Sinica, Taiwan

²UCLA, USA

³National Taiwan University, Taiwan

Motivation

- Investigate **fundamental** properties of opportunistic networks
- Better understand **network connectivity**
- Solve the long been ignored **censorship issue**

Contribution

- Point out and recover censorship within mobility traces of opportunistic networks
 - Propose **Censorship Removal Algorithm**
 - Recover censored measurements
- Prove the inter-contact time process as **self-similar** for future research on opportunistic networks

Outline

- Trace Description
- Censorship Issue
 - Survival Analysis
 - Censorship Removal Algorithm
- Self-similarity

Trace Description

- *UCSD campus trace**
 - 77 days, 275 nodes involved
 - Client-based trace
 - PDAs record Wi-Fi based APs nearby
- *Dartmouth College trace***
 - 1,777 days, 5148 nodes involved
 - Interface-based trace
 - APs maintain the association log for each wireless interface
 - 77 days extracted for comparison

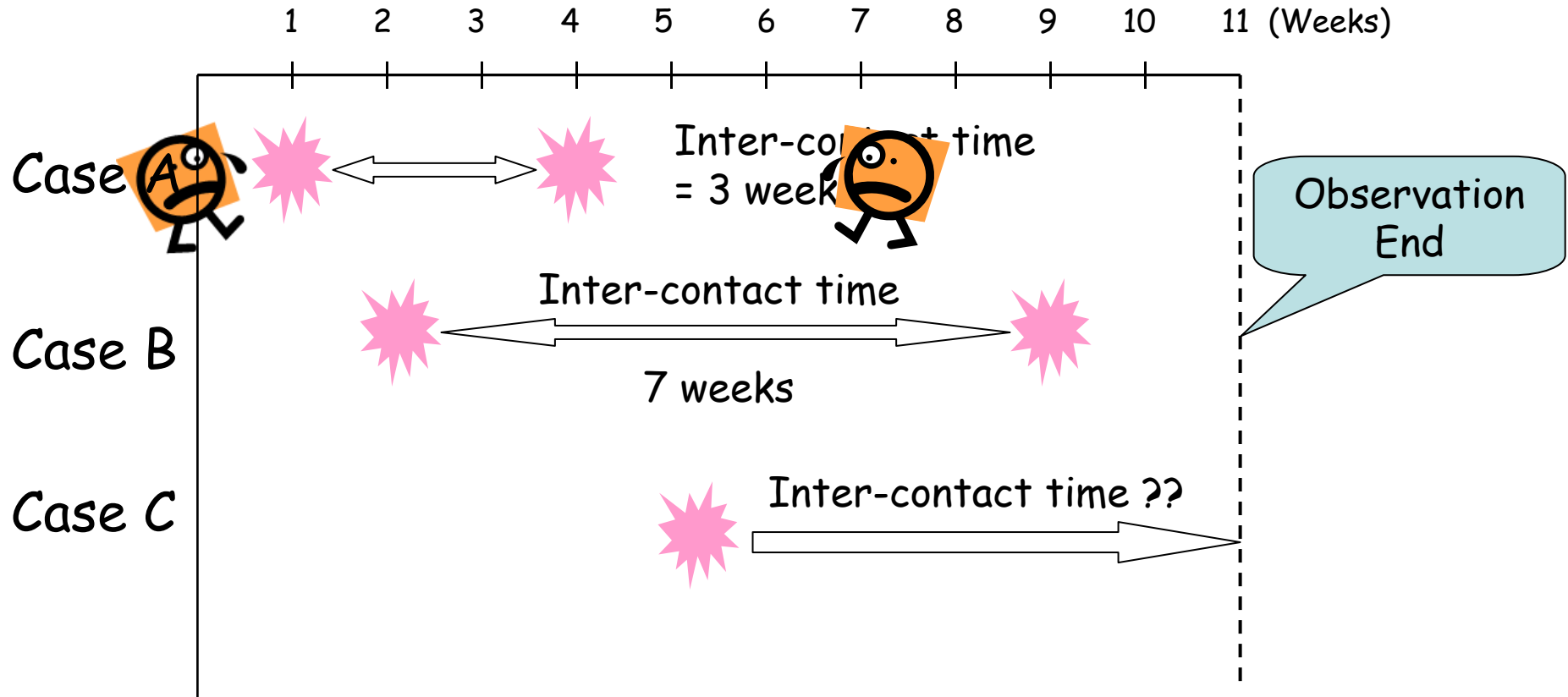
*UCSD: *Wireless Topology Discovery (WTD Project)*

**Dartmouth:  *RAW DAD*

Basic Terms

- What is **Contact** ?
 - Two nodes are of their wireless radio range
 - Associated to the same AP at the same time
- What is **Inter-contact Time** ?
 - Period between two consecutive contacts
- Used to observe **Network Connectivity**
 - Distribution of inter-contact time
 - Disconnection duration
 - Reconnection frequency

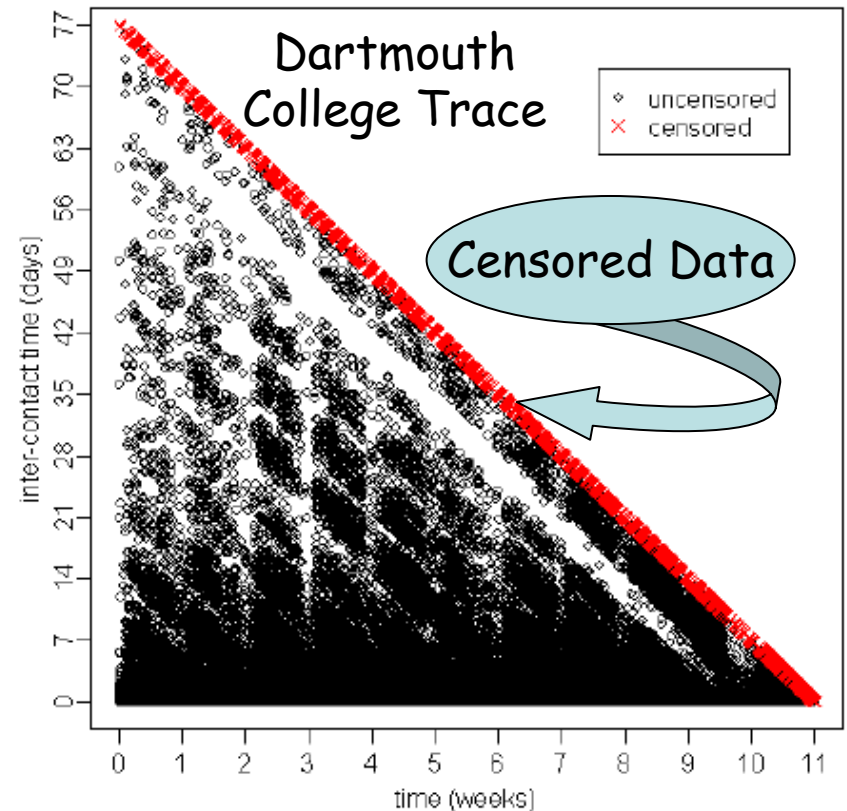
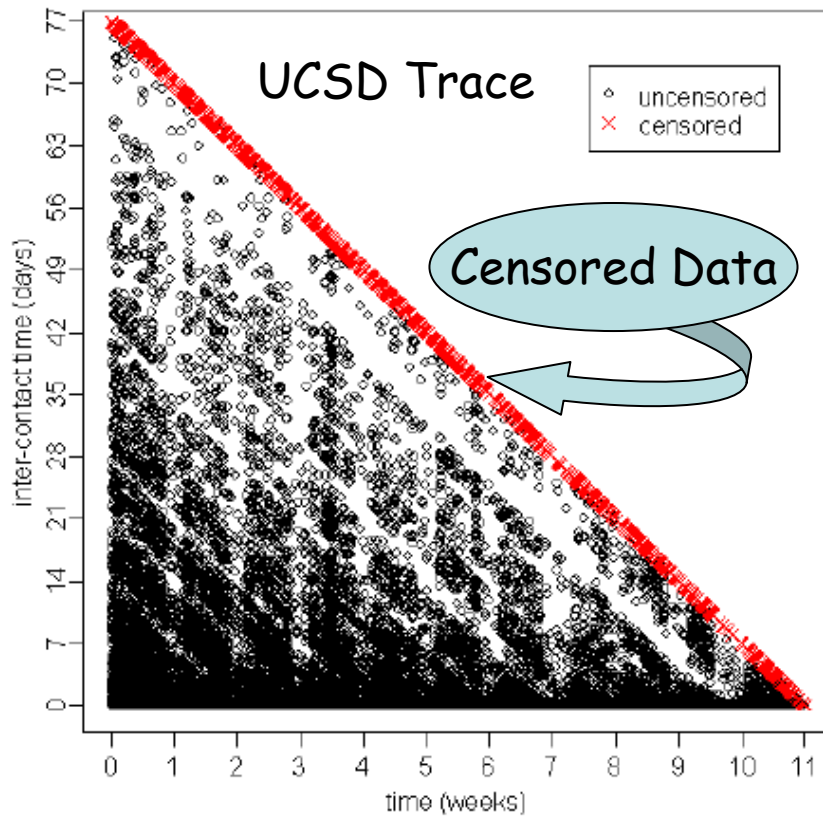
Basic Terms (Con't)



- In the last case, the inter-contact time has been **censored** as 6 weeks.

Censorship

- Inter-contact time samples end after the termination of the observation.
- Censored measurements are inevitable.



Survival Analysis

- Important in biostatistics, medicine, ...
 - Estimate patients' **time to live/death**
 - Map to censored inter-contact time samples
- Censored samples should have the same **likelihood distribution** as the uncensored's.
 - *Kaplan-Meier* Estimator
(a.k.a. *Survival Function* or *Product Limit Estimator*)

Kaplan-Meier Estimator

- Suppose there are N samples ($t_1 < t_2 < t_3 \dots < t_N$)
- At time t_i :
 - d_i uncensored samples (complete samples)
 - n_i events (censored/uncensored)
- The survival function is:

$$\begin{aligned}\hat{S}(t) &= \prod_{t_i \leq t} \Pr [t > t_i | t \geq t_i] \\ &= \begin{cases} 1 & ; t_1 > t \\ \prod_{t_i \leq t \leq t_N} \left[\frac{n_i - d_i}{n_i} \right] & ; t_1 \leq t \end{cases}\end{aligned}$$

Kaplan-Meier Estimator - An Example

- 10 inter-contact time samples:
 $1, 2^+, 3^+, 3.5^+, 4, 5^+, 9, 9.5^+, 10, 11^+$
 (in weeks, + for censorship)

i-c time interval	n_i	d_i (death)	c_i (censored)	Survival function $S(t)$
0	10	0	0	$S(0)=1$
(0,1]	10	1	0	$S(1)= 1 * 9/10=0.9$
(1,4]	6	1	3	$S(4)=0.9 * 5/6=0.75$
(4,9]	4	1	1	$S(9)=0.75 * 3/4=0.56$
(9,10]	2	1	1	$S(10)=0.56 * 1/2=0.28$
(10,11]	1	0	1	$S(11)=0.28 * 1/1= 0.28$

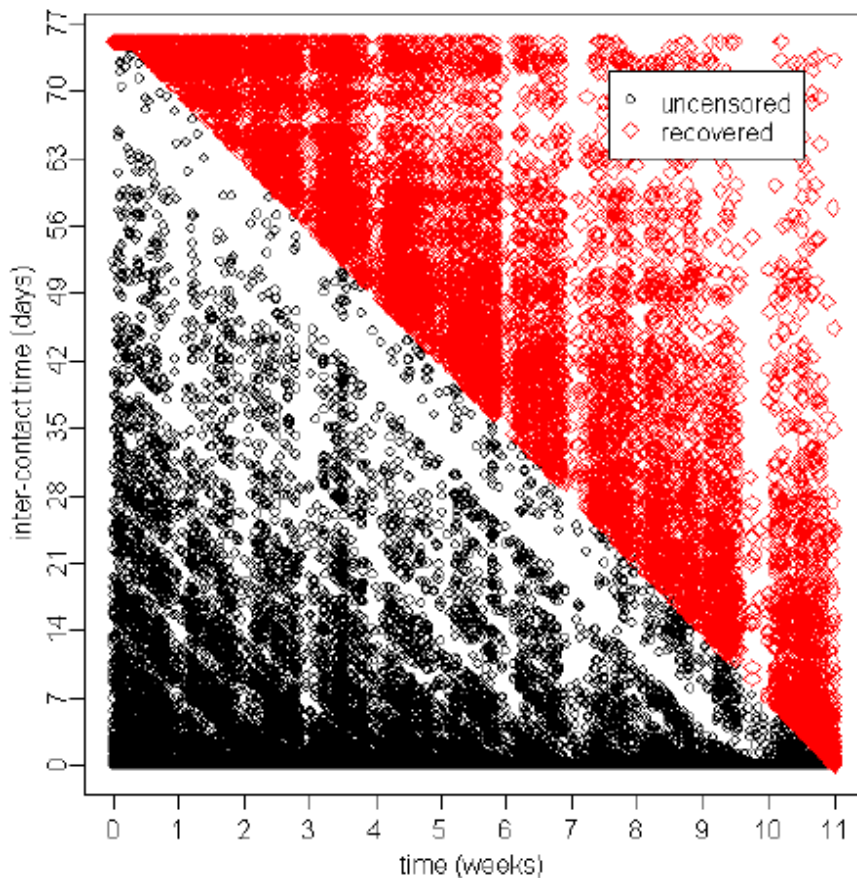
Censorship Removal Algorithm

- Based on the survival function $S(t)$
 - $t_1 < t_2 < t_3 \dots < t_N$ (N : total sample number)
 - **Death Ratio** during $t_i \sim t_{i+1}$: $D(t_i) = \frac{S(t_{i-1}) - S(t_i)}{S(t_i)}$
 - C_i : # of censored samples at t_i
 - Iteratively select $C_i * D(t_i)$ samples from C_i
 - Uniformly distribute their estimated inter-contact time by $S(t_i)$
 - Mark them as uncensored samples
 - Terminate when all the censored samples are removed

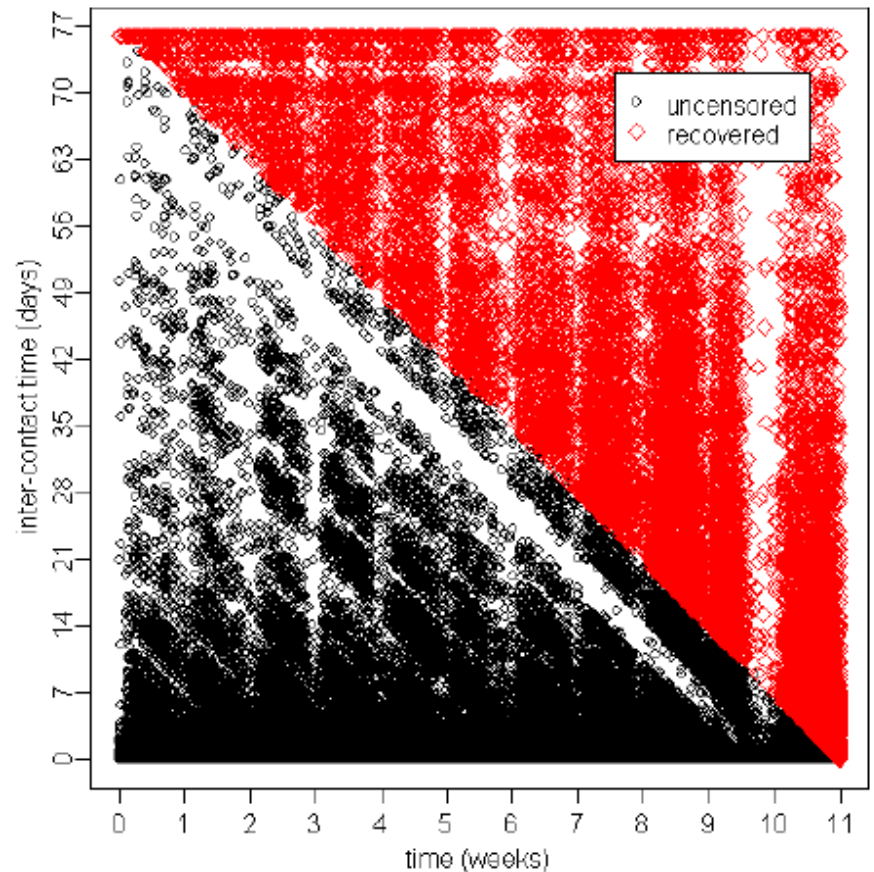
Censorship Removal Algorithm (Con't)

- Recovered inter-contact time measurements

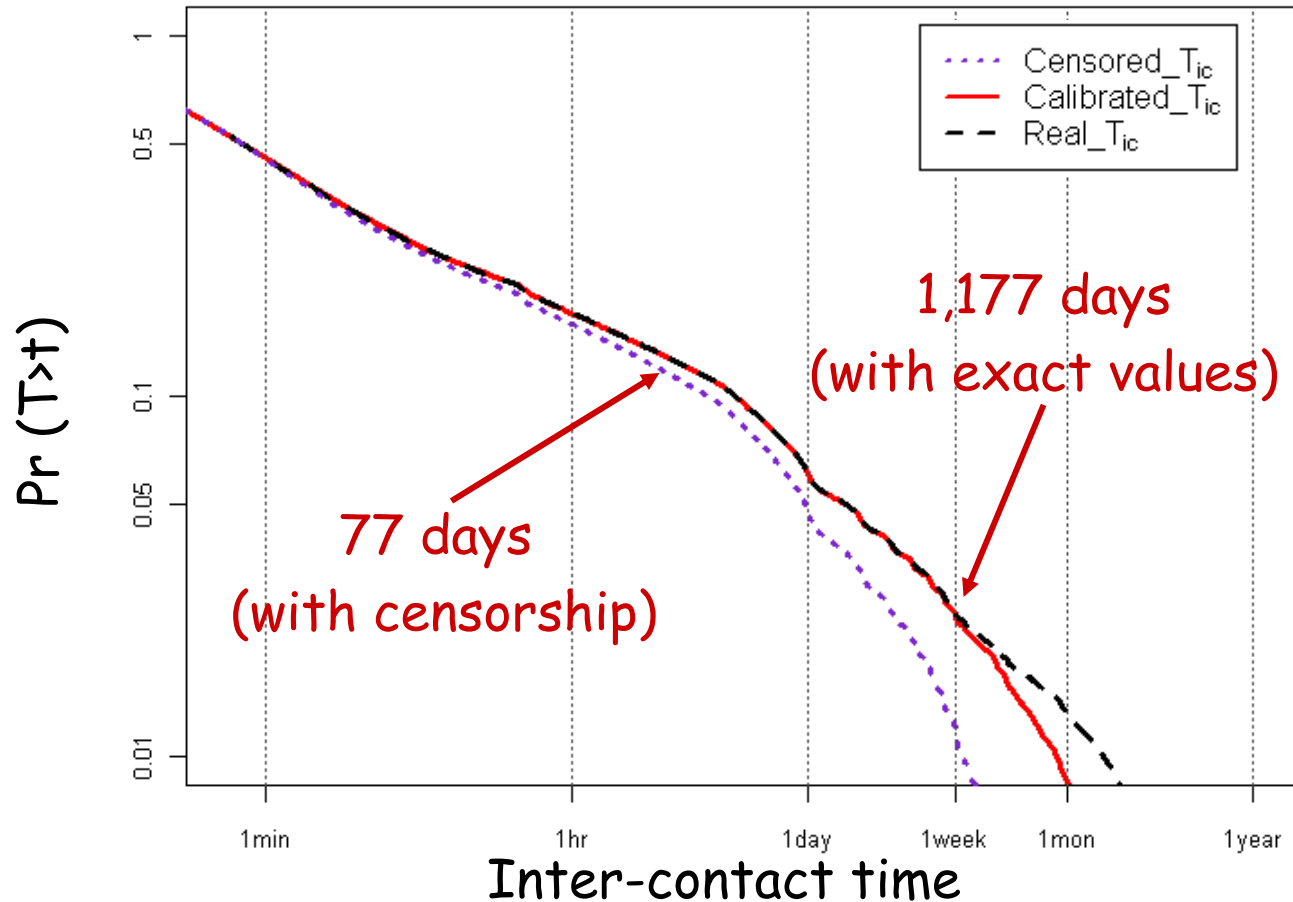
UCSD Trace



Dartmouth Trace



Censorship Removal Algorithm (Con't)



- Compare the **recovered values** to their **exact values** in original trace.
- 80.4% censored measurements are recovered.

Outline

- Trace Description
- Censorship Issue
 - Survival Analysis
 - Censorship Removal Algorithm
- *Self-similarity*

Self-Similarity

- What is self-similarity?
 - By definition, a self-similar object is **exactly** or approximately **similar** to part of itself.
- In opportunistic network, we focus on the **network connectivity**
- With recovered measurements, we prove **inter-contact time series** as a self-similar process
 - Reconnection/disconnection
 - Similar mobility pattern in people opp. networks

Self-Similarity

- A self-similar series
 - Distribution should be heavy-tailed
 - Examined by three statistical analyses
 - Variance-Time Plot, R/S Plot, Periodogram Plot
 - Estimated by a specific parameter : *Hurst*
 - H should be in the range of $0.5 \sim 1$
 - Results of three methods should be in the 95% confidence interval of Whittle estimator

Self-Similarity (Con't)

- Previous works show inter-contact time dist. as power-law dist.
- A random variable X is called heavy-tailed:
 - If $P[X > x] \sim cx^{-\alpha}$, with $0 < \alpha < 2$ as $x \rightarrow \infty$
 - α can be found by log-log plot
 - Survival curves show the α for
 - UCSD: 0.26
 - Dartmouth: 0.47

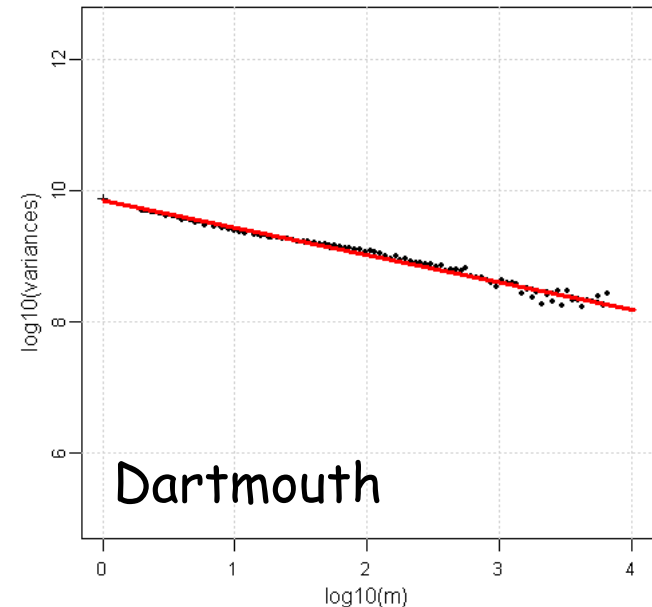
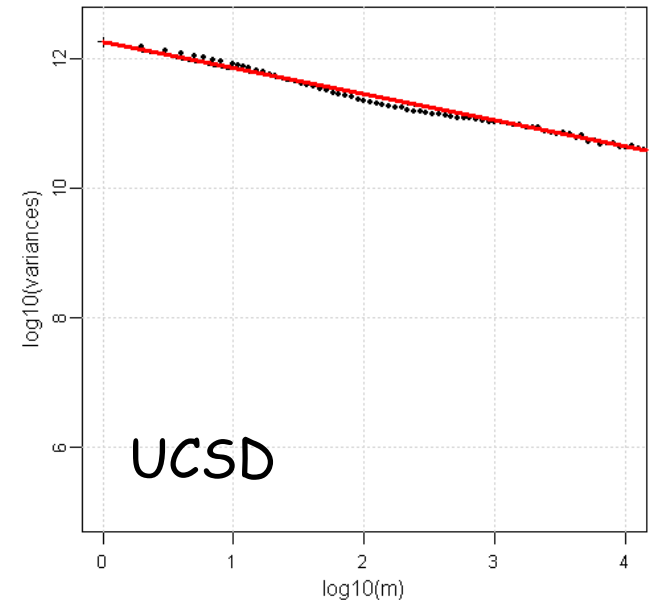
Self-Similarity (Con't)

- *Variance-Time Method*

- Variance decreases very slowly, even when the size grows large

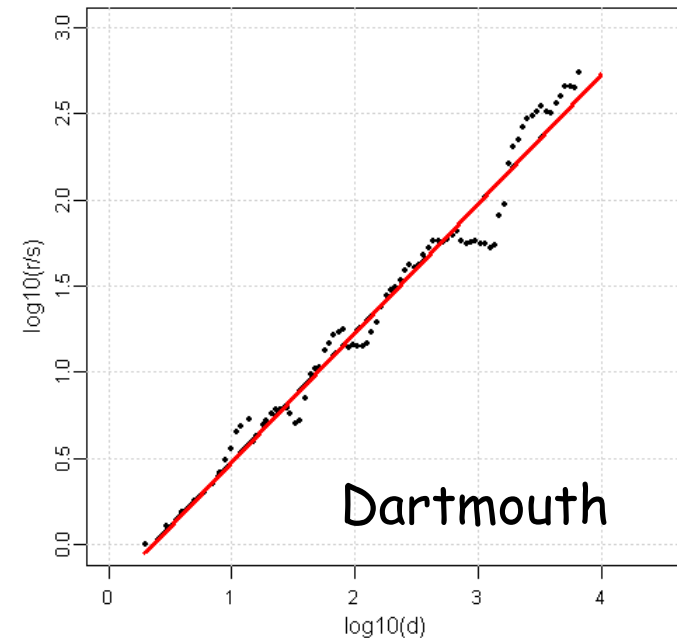
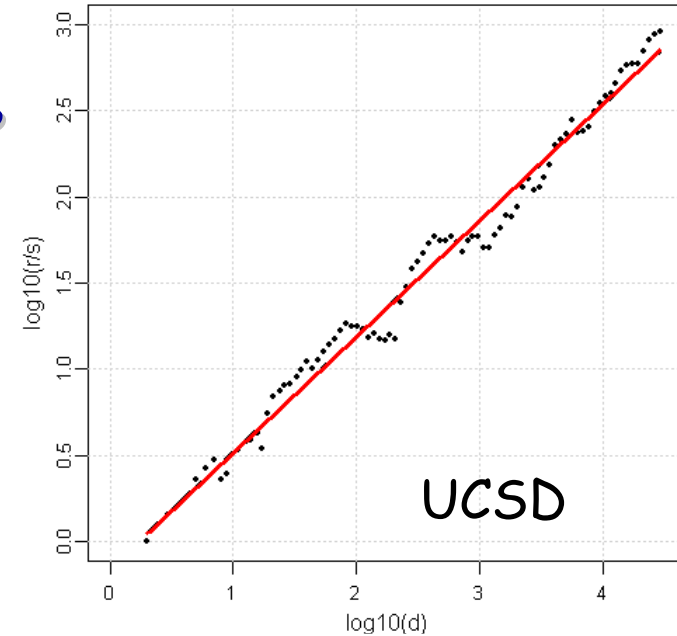
- The Hurst estimates are

- UCSD: 0.801
 - Dartmouth: 0.7973



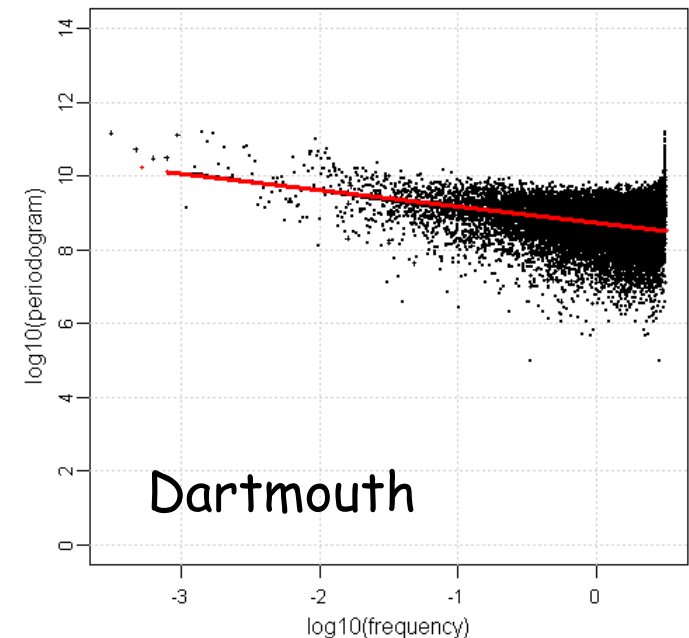
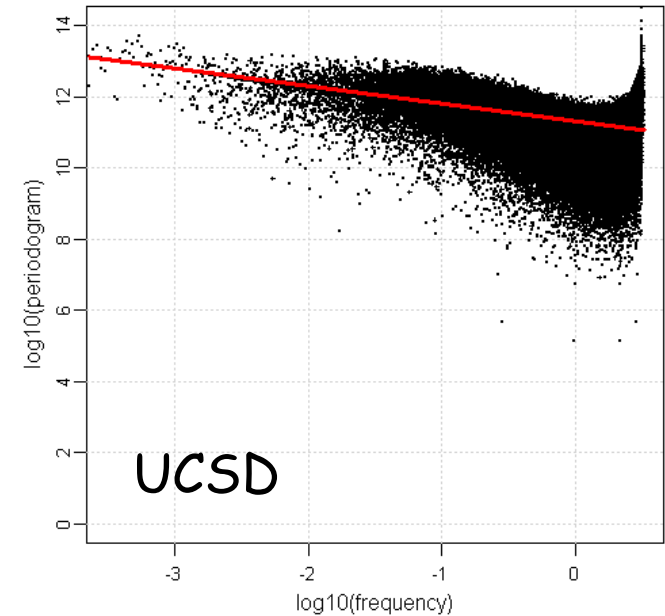
Self-Similarity (Con't)

- *Rescaled Adjusted Range (R/S) method*
 - Keep similar properties when the dataset is divided into several sub-sets
- The Hurst estimates are
 - UCSD:0.7472
 - Dartmouth:0.7493

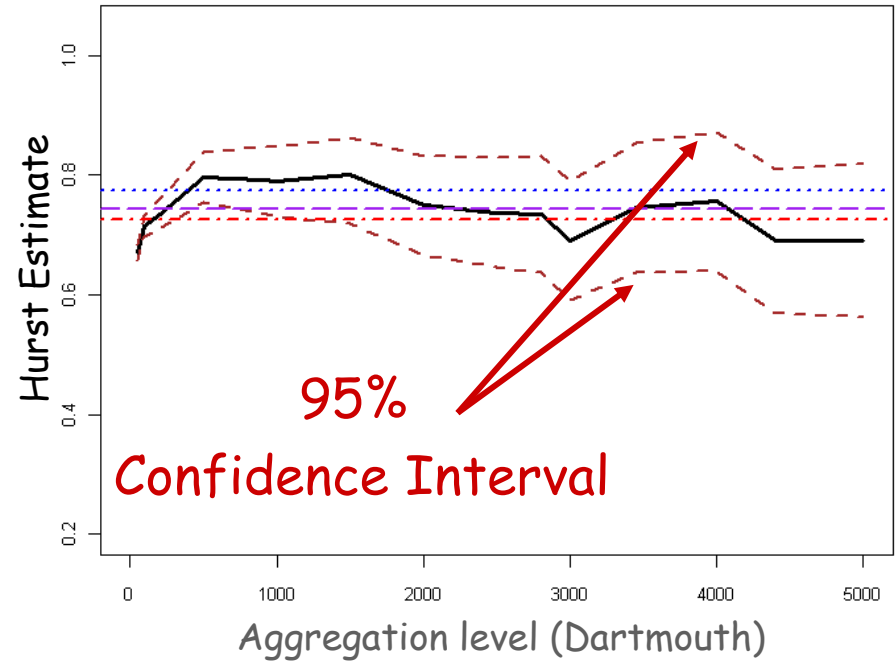
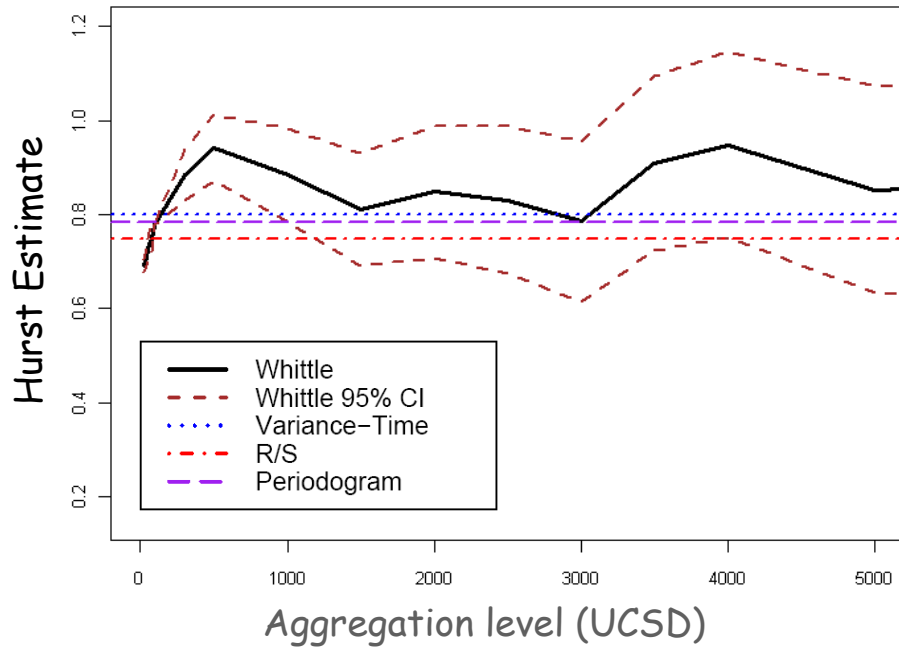


Self-Similarity (Con't)

- *Periodogram Method*
 - Use the slope of **power spectrum** of the series as frequency approaches zero
- The Hurst estimates are
 - UCSD: 0.7924
 - Dartmouth: 0.7655



Self-Similarity (Con't)



- **Whittle Estimator**
 - Usually being considered as a **more robust** method
 - Provide a confidence interval
- Results of the three graphical methods are in the **95% confidence interval**.

Conclusion

- Two major properties exist in modern opportunistic networks:
 - Censorship
 - Self-similarity
- Using CRA, we could recover censored inter-contact time to have more accurate datasets.
- With recovered datasets, we prove that inter-contact time series is self-similar.

Thank You !