Diffusion, Persistence and Mobility: An Overview

Omar Malik
Overview

- **Diffusion and Persistence**
  - The diffusion equation
  - Persistence as a measure of diffusion
  - Persistence of different networks
  - What can it be used for?

- **Mobility**
  - Human mobility: predictable and cyclical
  - Mobility patterns and (the lack of) anonymity
  - Mobility patterns from scarce data
Diffusion and Persistence
Diffusion

- The net statistical movement of particles.
- Described by a differential equation.
- The equation may be discretized.

\[ u_{i+1}^{n+1} = u_i^n + F(u_{i+1}^n - 2u_i^n + u_{i-1}^n) . \]
Persistence

- Persistence refers to whether or not a node changes state up to a given time.
- When studying diffusive processes persistence becomes a measure of the progress of diffusion.
Regular Networks
1-D Lattices
Regular Networks
2-D Lattices

Diffusive persistence on 2D disordered lattices ($n = 10^4$)
Random Networks
ER Networks
Persistence

- Network topology determines the persistence behavior
  - Networks with regular structure show power-law scaling.
  - This scaling is absent in random graphs.
- What good is it?
  - Indicates how susceptible a network is to any process that can be modelled diffusively.
Mobility
Human mobility: random or predictable?

Older models tended to model human mobility patterns with stochastic processes.

However, Barabasi showed that while randomness exists, most people spend > 60% of their time in just two locations.
Mobility

- Cell phone data was used to track user’s locations.
- Entropy can be used to characterize the amount of randomness in a user’s mobility patterns.
- Users’ locations could potentially be predicted with 93% accuracy.
- Independent of the radius of gyration
Mobility

Graph A: Distribution of entropy $P(S)$ for different types of mobility models.

Graph B: Distribution of entropy $P(\Pi)$ for different types of mobility models.

Graph C: Maximum entropy $\Pi_{\text{max}}$ as a function of range $r_g$ (km).

Graph D: Cumulative distribution of top locations visited $\xi$ as a function of $n$.
Anonymity

- Just four randomly chosen points are enough to uniquely identify an individual user with ~95% certainty.
Anonymity

- Even with loss of resolution the pattern remains highly predictive.
Sparse Self-submitted Data

- So far the data we have been talking about is gathered automatically.
- If the data is user-submitted (and therefore exceptional rather than typical) does it exhibit more randomness?
- Even in self-submitted data the top two locations represent ~60% of the check-ins.
Sparse Self-submitted Data

- By using clustering techniques we can still determine behavioral patterns.
So far the data we have been talking about is gathered automatically. If the data is user-submitted (and therefore exceptional rather than typical) does it exhibit more randomness? Even in self-submitted data the top two locations represent ~60% of the check-ins.
Questions