Data Engineering Research Group

2 faculty members

Reynold Cheng

Ben Kao

20 research students (10 PhD, 10 MPhil)

Success Stories

Papers in the past 5 years

- 58 in top DB and DM conferences (9 SIGMOD, 15 VLDB, 17 ICDE, 3 EDBT, 4 CIKM, 3 SIGKDD, 6 ICDM)
- 28 in top DB and DM journals (5 TODS, 7 VLDBJ, 15 TKDE, 1 TKDD)

PhD alumni with faculty positions

(Rutgers, HKPolyU, Mexico State U, Aalborg U, Macau U, Renmin U)

Reynold Cheng

Background

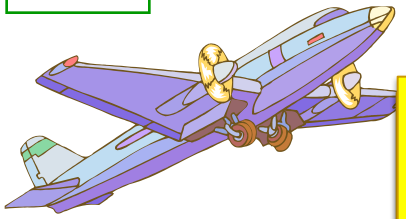
- HKU (BSc, MPhil 95-00), Purdue (PhD, 00-05), HKPolyU (Asst. Prof, 05-08)

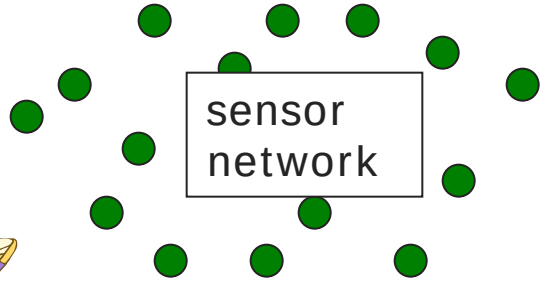
Research

- Database management, uncertainty management, data mining, spatial databases

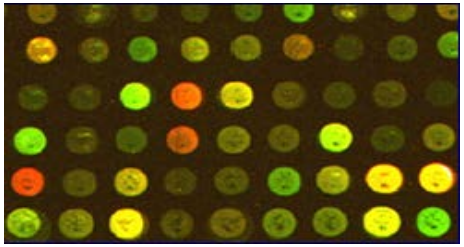
Data Uncertainty

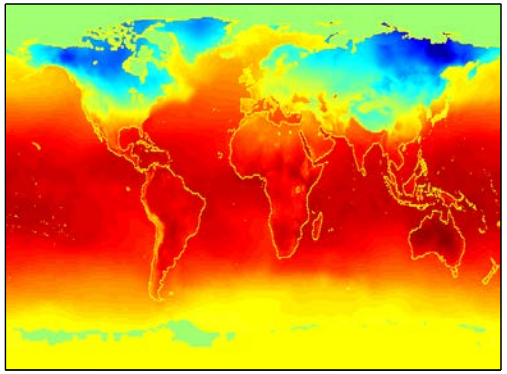
GPS





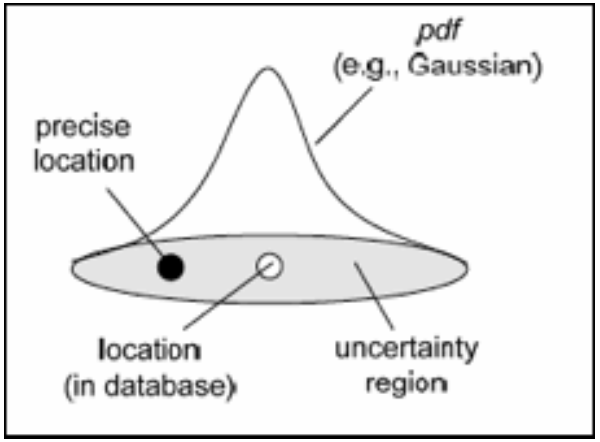
Handle data uncertainty, or service quality can be degraded!

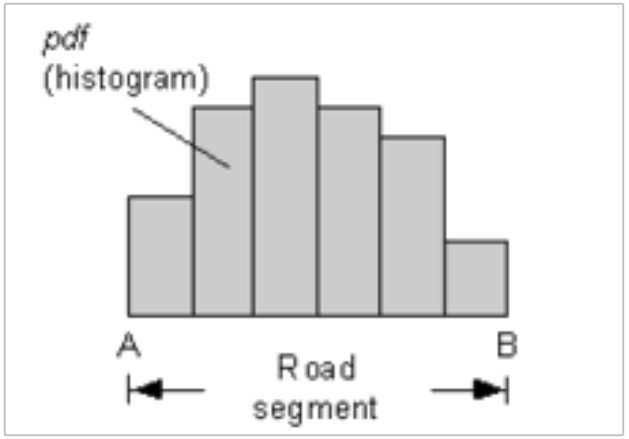




The ORION Database

- Treat data uncertainty as a first-class citizen





- A probabilistic query provides answers with probabilities (e.g., Mary has a 80% chance to be in HKU)

Queries in ORION

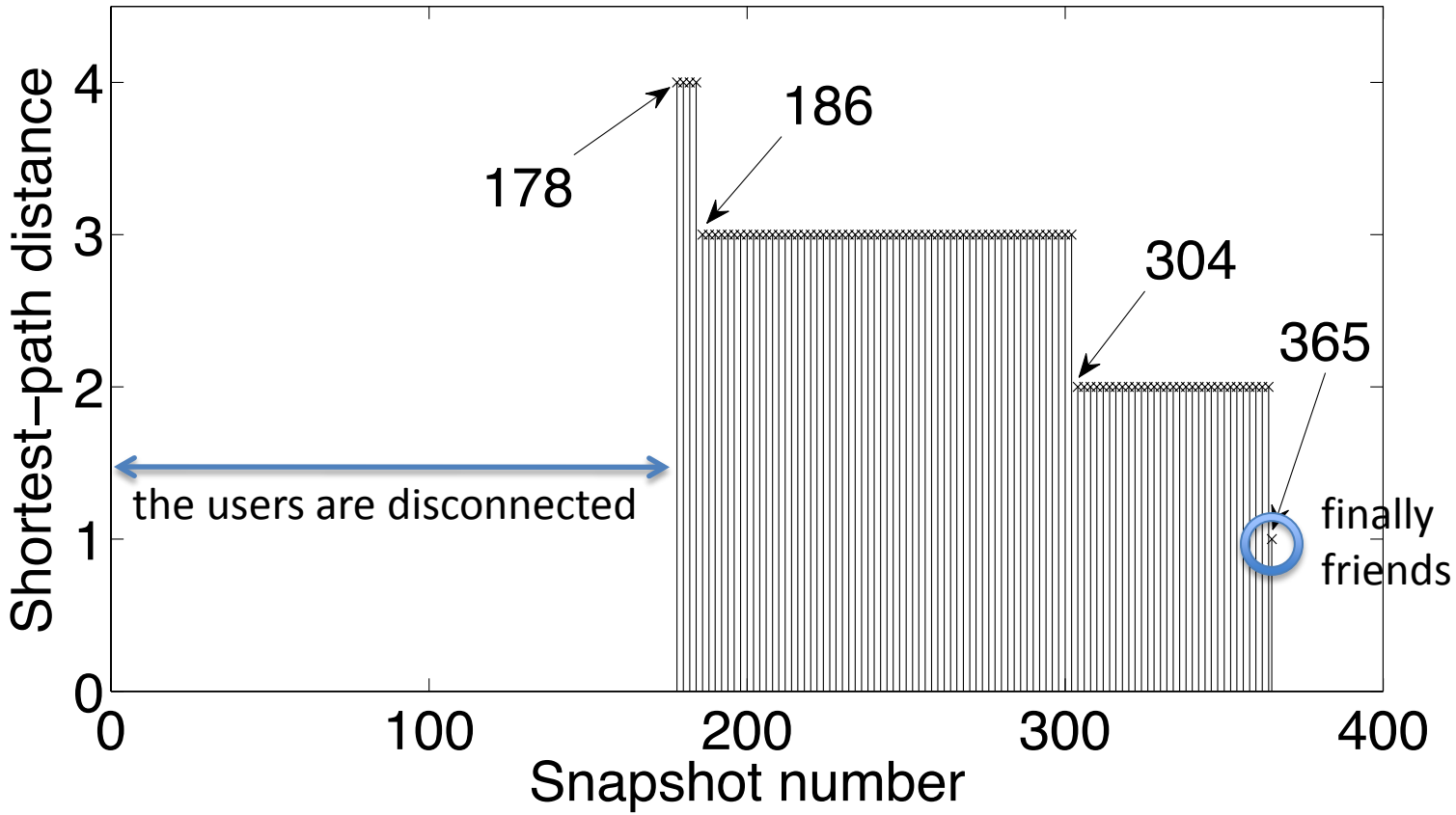
k	*a*
1	*U*[5,10]

Create a table with **UNCERTAIN** type	`CREATE table T( k INTEGER primary key, a UNCERTAIN);`
Insert Gaussian pdf (μ,σ)	`Insert into T values (2,'(g,μ,σ)');`
Display uncertain info. of *a* if *a* > 5	`SELECT a FROM T where a > 5;`
Equality join of uncertain attributes (=% returns probability of equality)	`SELECT R.k, S.k, R.a =% S.a FROM R,S WHERE R.a = S.a;`
Entities with prob. giving min value of *a* (e.g., {(3,0.5), (5,0.3), (11,0.2)})	`SELECT Emin(T.a) from T;`
Min value of *a* for table *T* (UNCERTAIN)	`SELECT Vmin(T.a) from T;`

Ben Kao

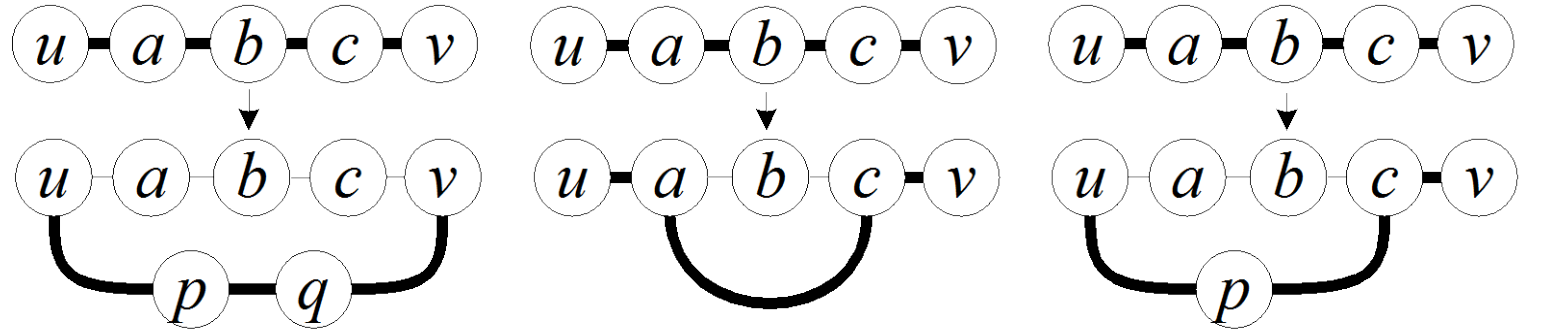
- Background
  - HKU (BSc 86-89), Princeton-Stanford (PhD 89-95)
- Research
  - Database Systems, Information Retrieval, Data Mining

Finding Key Moments in Social Networks



"Distance" between two Facebook users over a 1-year period. (They are disconnected before Day 178 and finally became friends on Day 365.)

How did they (*u* and *v*) become friends?



- To understand how friendships are established, we need to study the events that happened at certain "key moments". For example, what happened (Events (a), (b) or (c) above) that led to the shortening of two users' distance from each other?
- But first, we need to discover those "key moments" so we know which "snapshots" of the Facebook graph we should look at.

Evolving Graph Sequence (EGS) Processing

- We model the dynamics of a social network as a (big) sequence of (big) evolving graph snapshots.
- We study efficient graph algorithms for identifying key moments (snapshots at which sharp changes in certain key measures are observed).
- Such key moments help social network analysts investigate the various properties of gigantic social networks.