



Week 9 – Sensor and dataset processing

Lovett et al. – “The Calendar as a Sensor: Analysis and Improvement Using Data Fusion with Social Networks and Locations”

Adam Goldstein

abg@wpi.edu

CS 525w – Mobile Computing (03/22/2011)





The Calendar as a Sensor

- Hypothetically, calendars contain information about when people are where
- In reality, this 1:1 relationship does not exist.
 - Not all event attenders actually attend
 - Not all events are even real events
- Here are 4 pages about why calendars are inaccurate.



Preliminary Study

- Authors started with a 6 week study
 - About 200 software developers and engineers
 - Actually included about 20 employees
 - Used MS Outlook
 - Events were mined from participants calendars
 - Actual activity was discovered by observation, interviews, and participant diaries.



Event Types (474 total)

- Genuine events (38 total, 8%)
 - Shared cal event with ≥ 1 participant
- Placeholder events (152, 32%)
 - ≥ 1 participant but doesn't actually occur, e.g. Repeated meeting that was cancelled
- Personal reminders (232, 49%)
 - Not a physical event, but a note by one person to themselves
- Shared reminders (52, 11%)
 - Not a physical event, but a note to many people



Expectation vs. Reality

- Comparing real world attenders and invited-on-the-calendar attenders

$$\overline{J(V_r, V_c)} = \frac{1}{N} \sum_{j=1}^N \left[\frac{|V_r \cap V_c|}{|V_r \cup V_c|} \right]_j$$

- In other words:

$$\left(\sum_{\text{All events}} \frac{[\text{Invited people who actually came}]}{[\text{Invited people and attending people}]} \right)$$

[All events]



Expectation vs. Reality Results

Comparison	Real world
Start time (nearest 5 minutes)	(-25, 25)
End time (nearest 5 minutes)	(-5, 15)
Location	0.11
Total correct identification	113
Total false identification	16
Total failed identifications	9
Set similarity (Jaccardian index)	0.89

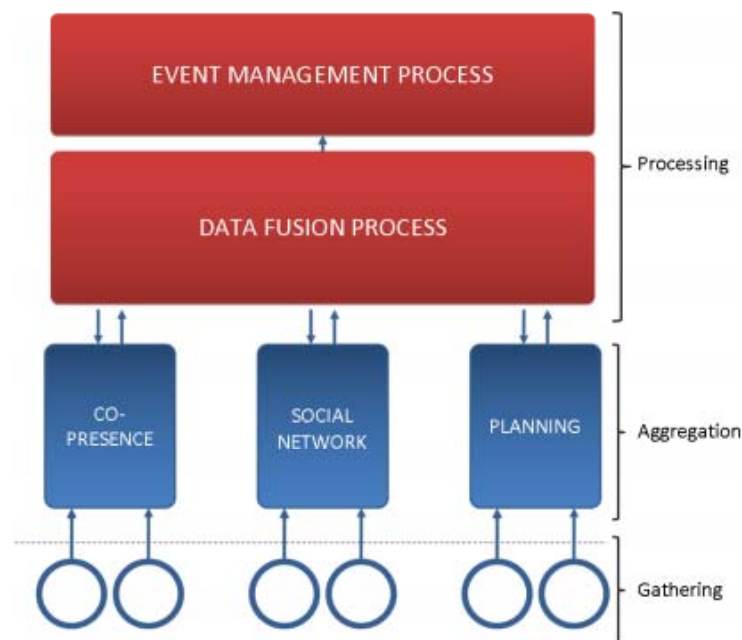


So we can agree

- The calendar alone is not a sufficient sensor.
- What if we fused it with not one, not two, but three other data sources?
 - You must have read the rest of this paper

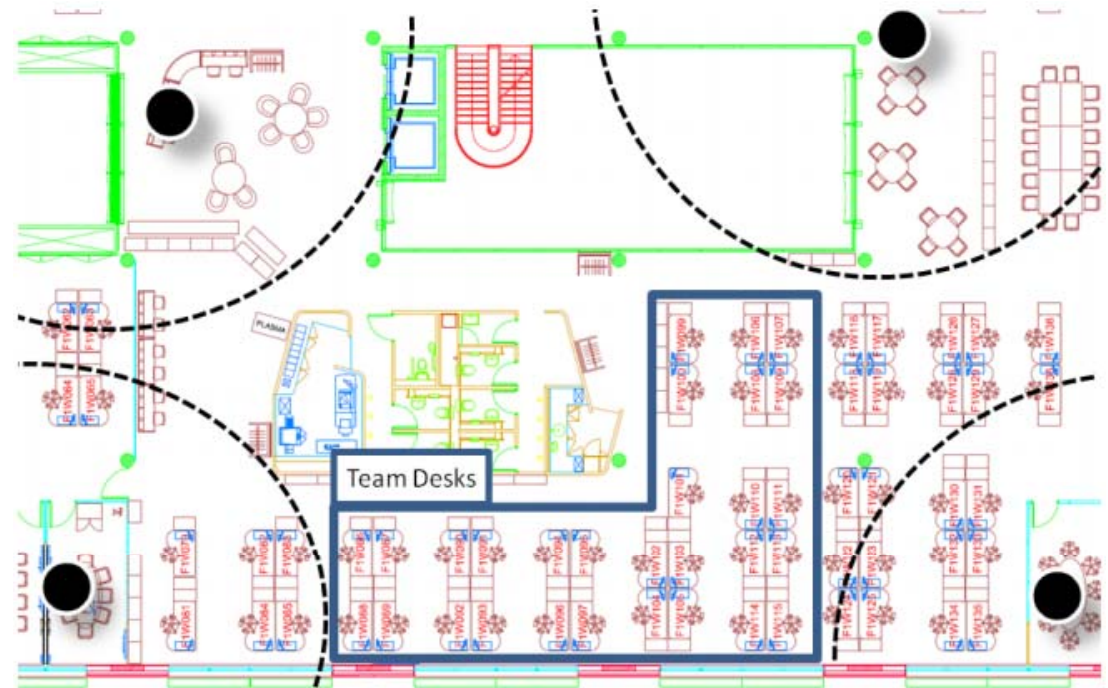
Data fusion - Enablers

- We can make the calendar a viable sensor by adding three other information sources:
 - Co-presence
 - “Social network”
 - Planning



Co-location

Gather all users who are in the same proximity (within threshold p) for a decent amount of time (within threshold t).



$$U_c = \{u : u \in U \wedge |l - l_u| \leq p \wedge (t_m - t_u) \leq t\}$$



“Social Network”

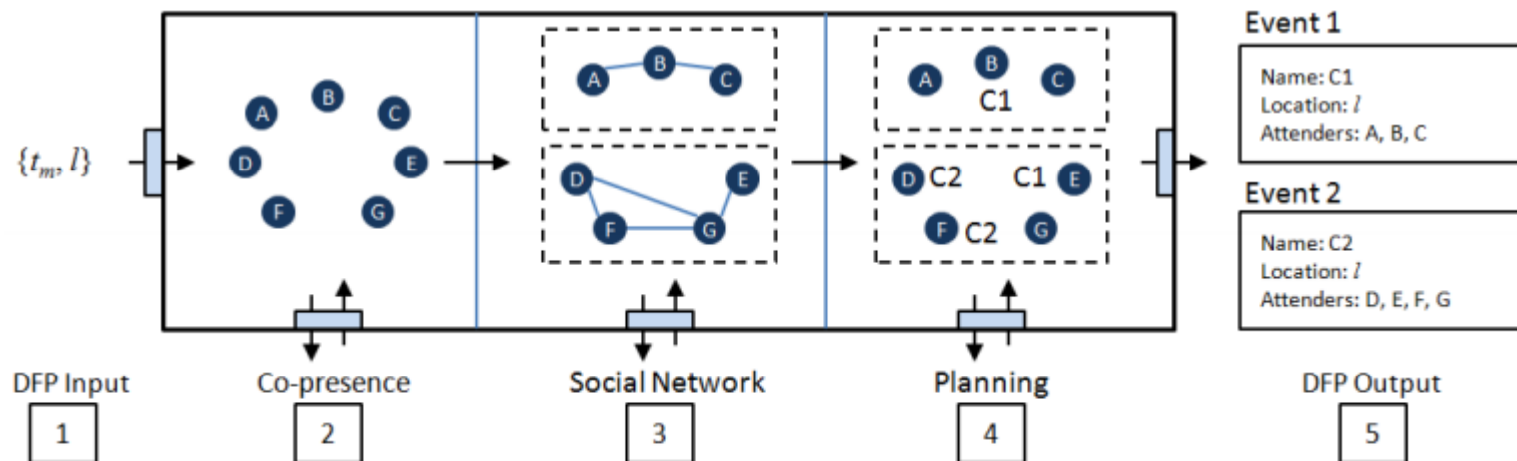
- If given a set of users
 - Create graphs of all users who have social ties (in each others’ contact list)
 - Graphs are ≥ 2 people
- If given graphs of users with events
 - Attach ungrouped users into graphs based on social ties



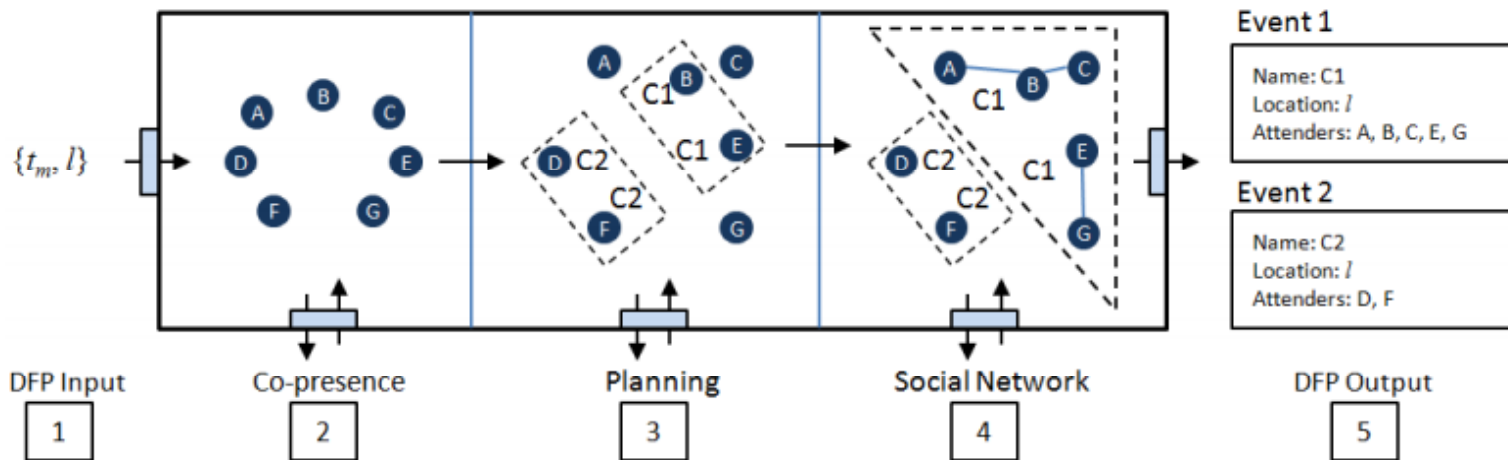
Planning

- If given a set of users
 - Create graphs of all users who have shared events together at a specific time
- If given graphs of users in the same “social network”
 - Reshape the graphs according to scheduled events that are most shared

Data fusion (Method 1)



Data fusion (Method 2)





Event Management Process

- When triggered, use data fusion to
 - Create events
 - Update events
 - End events
- This is where the real success came in

Metric	Original	Method 1	Method 2
Success event ID	38	37	32
False event ID	204	32	14
Failed event ID	N/A	1	6

14



Expectation vs. Reality Results

Comparison	Real world	Method 1	Method 2
Start time (nearest 5 minutes)	(-25, 25)	(-5, 20)	(0, 15)
End time (nearest 5 minutes)	(-5, 15)	(-5, 20)	(-5, 20)
Location	0.11	0.97	0.84
Total correct identification	113	112	94
Total false identification	16	36	31
Total failed identifications	9	10	29
Set similarity (Jaccardian index)	0.89	0.65	0.60



Thoughts & concerns

- False identifications
 - Privacy concerns, spam
- Failed identifications
 - Unreliable system
- Sensor failure
- Participant Mobility
 - Just passing by, using conference rooms for other reasons
- Anything else?



Conclusions

- With data fusion, calendars can be made a more genuine source of information
 - Number of false events improved from 204 using just a calendar to < 32 .
 - Updated calendars distinguish between genuine events are reminders
- Any other useful contributions?
- Other ways this fusion approach can be used?