Learning Objectives

- After this segment, students will be able to
  - Contrast colocations and associations
  - Describe colocation interest measures

Background: Association Rules

- Association rule e.g. (Diaper in T => Beer in T)

Transaction	Items Bought
1	{socks, , milk, , beef, egg, …}
2	{pillow, , toothbrush, ice-cream, muffin, …}
3	{ , , pacifier, formula, blanket, …}
…	…
n	{battery, juice, beef, egg, chicken, …}

  - Support: probability (Diaper and Beer in T) = 2/5
  - Confidence: probability (Beer in T \| Diaper in T) = 2/2

- Apriori Algorithm
  - Support based pruning using monotonicity

Apriori Algorithm

How to eliminate infrequent item-sets as soon as possible?

Transaction Id	Time	Item-types bought
1101	18:35	Milk, bread, cookies, juice
792	19:38	Milk, juice
2130	20:05	Milk, eggs
1735	20:40	Bread, cookies, coffee

Support threshold >= 0.5

Apriori Algorithm

Eliminate infrequent singleton sets

Transaction Id	Time	Item-types bought
1101	18:35	Milk, bread, cookies, juice
792	19:38	Milk, juice
2130	20:05	Milk, eggs
1735	20:40	Bread, cookies, coffee

Item-type	Count
Milk	3
Bread	2
Cookies	2
Juice	2
Coffee	1
Eggs	1

Support threshold >= 0.5

Milk Bread Cookies Juice Coffee Eggs

Apriori Algorithm

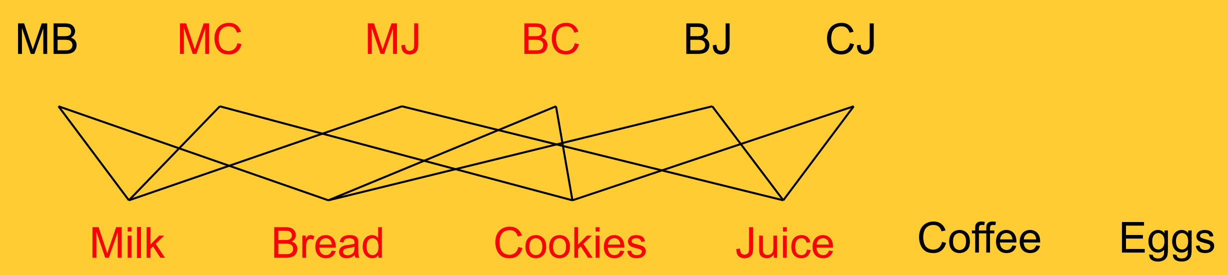
Make pairs from frequent items & prune infrequent pairs!

Transaction Id	Time	Item-types bought
1101	18:35	Milk, bread, cookies, juice
792	19:38	Milk, juice
2130	20:05	Milk, eggs
1735	20:40	Bread, cookies, coffee

Item-type	Count
Milk	3
Bread	2
Cookies	2
Juice	2
Coffee	1
Eggs	1

Item Pair	Count
Milk, Cookies	2
Milk, Juice	2
Bread, Cookies	2
Milk, Bread	1
Bread, Juice	1
Cookies, Juice	1

Support threshold >= 0.5



Apriori Algorithm

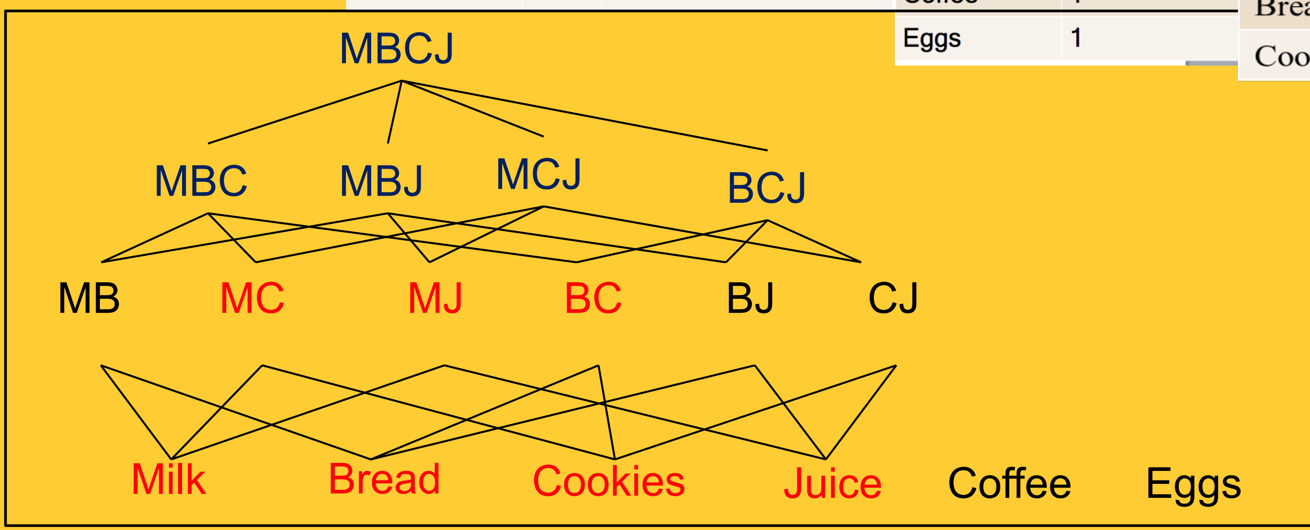
Make triples from frequent pairs & Prune infrequent triples!

Support threshold >= 0.5

Transaction Id	Time	Item-types bought
1101	18:35	Milk, bread, cookies, juice
792	19:38	Milk, juice
2130	20:05	Milk, eggs
1735	20:40	Bread, cookies, coffee

Item-type	Count
Milk	3
Bread	2
Cookies	2
Juice	2
Coffee	1
Eggs	1

Item Pair	Count
Milk, Cookies	2
Milk, Juice	2
Bread, Cookies	2
Milk, Bread	1
Bread, Juice	1
Cookies, Juice	1



No triples generated Due to Monotonicity!

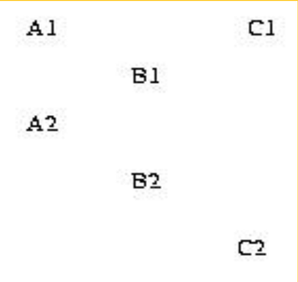
Apriori algorithm examined only 12 subsets instead of 64!

Association Rules Limitations

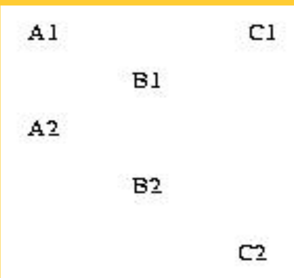
- **Transaction is a core concept!**
  - Support is defined using transactions
  - Apriori algorithm uses transaction based Support for pruning

Transaction	Items Bought
1	{socks, , milk, , beef, egg, …}
2	{pillow, , toothbrush, ice-cream, muffin, …}
3	{ , , pacifier, formula, blanket, …}
…	…

- However, spatial data is embedded in continuous space
  - Transactionizing continuous space is non-trivial !



Spatial Association vs. Cross-K Function



Input = Feature A,B, and, C, & instances A1, A2, B1, B2, C1, C2

- Spatial Association Rule (Han 95)
- Output = (B,C) with threshold 0.5
- Transactions by Reference feature, e.g. C

  Transactions: (C1, B1), (C2, B2)

  Support (A,B) = Ǿ

  Support(B,C)=2 / 2 = 1

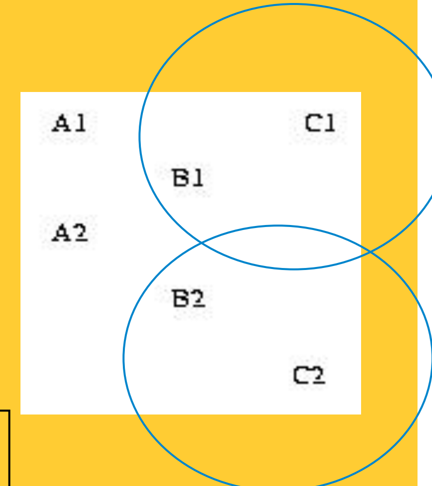
- Cross-K Function

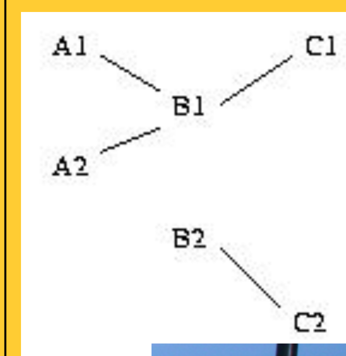
  Cross-K (A, B) = 2/4 = 0.5

  Cross-K(B, C) = 2/4 = 0.5

  Cross-K(A, C) = 0

  Output = (A,B), (B, C) with threshold 0.5





Spatial Colocation

Features: A. B. C

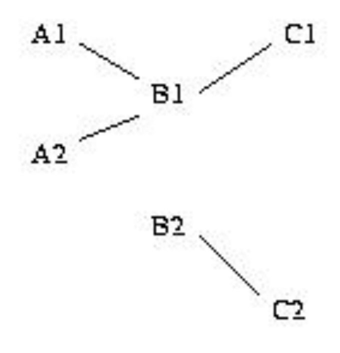
Feature Instances: A1, A2, B1, B2, C1, C2

Feature Subsets: (A,B), (A,C), (B,C), (A,B,C)

Participation ratio (pr):

pr(A, (A,B)) = fraction of A instances neighboring feature {B} = 2/2 = 1

pr(B, (A,B)) = ½ = 0.5



Participation index (A,B) = **pi**(A,B) = min{ **pr**(A, (A,B)), **pr**(B, (A,B)) } = min (1, ½) = 0.5

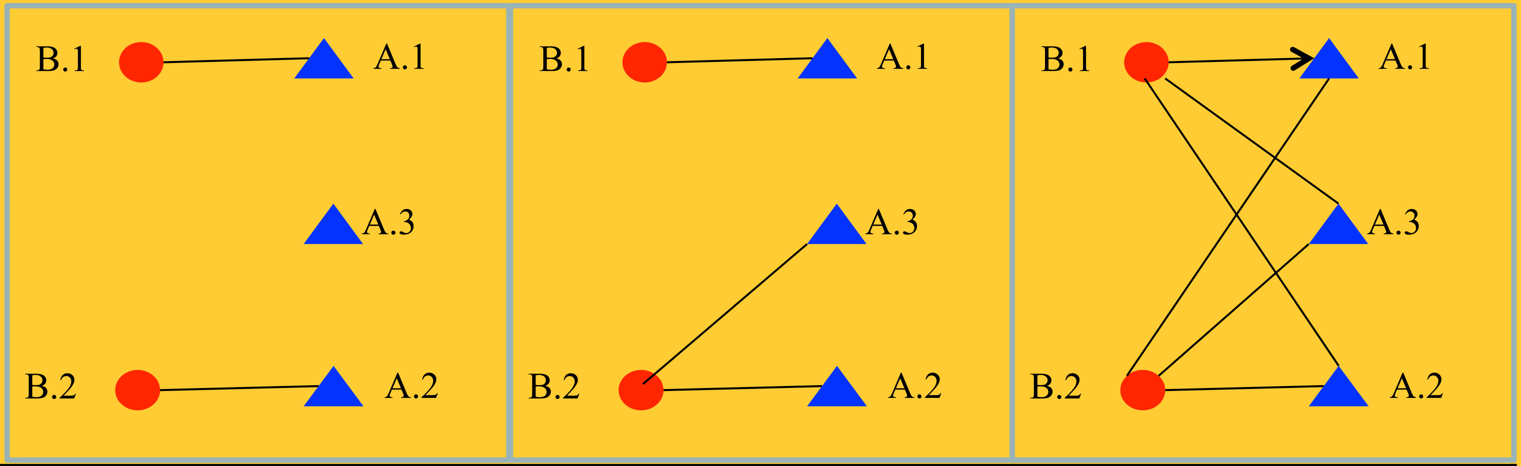
pi(B, C) = min{ **pr**(B, (B,C)), **pr**(C, (B,C)) } = min (1,1) = 1

Participation Index Properties:

(1) <u>Computational</u>: Non-monotonically decreasing like support measure

(2) <u>Statistical</u>: Upper bound on Ripley's Cross-K function

Participation Index >= Cross-K Function

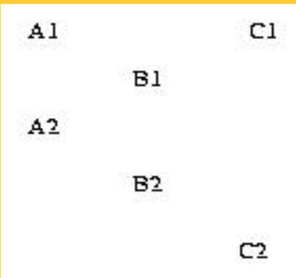


Cross-K (A,B)	2/6 = 0.33	3/6 =0.5	6/6 = 1
PI (A,B)	2/3 = 0.66	1	1

Association Vs. Colocation

	Associations	Colocations
underlying space	Discrete market baskets	Continuous geography
event-types	item-types, e.g., Beer	Boolean spatial event-types
collections	Transaction (T)	Neighborhood N(L) of location L
prevalence measure	Support, e.g., Pr.[Beer in T]	Participation index, a lower bound on Pr.[A in N(L) | B at L]
conditional probability measure	Pr.[Beer in T | Diaper in T]	Participation Ratio(A, (A,B)) = Pr.[A in N(L) | B at L]

Spatial Association Rule vs. Colocation



Input = Spatial feature A,B, C, & their instances

- Spatial Association Rule (Han 95)
- Output = (B,C)
- Transactions by Reference feature C

  Transactions: (C1, B1), (C2, B2)

  Support (A,B) = Ǿ, Support(B,C)=2 / 2 = 1

- Cross-K Function

  Cross-K (A, B) = 2/4 = 0.5

  Cross-K(B, C) = 2/4 = 0.5

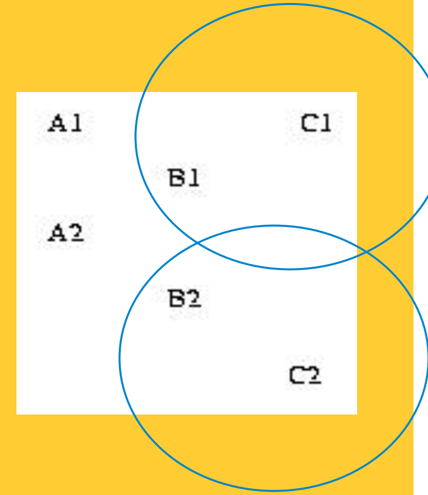
  Output = (A,B), (B, C)

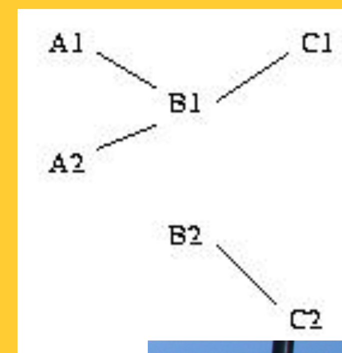
- Colocation - Neighborhood graph

  Output = (A,B), (B, C)

  PI(A,B) = min(2/2,1/2) = 0.5

  PI(B,C) = min(2/2,2/2) = 1





Spatial Colocation: Trends

- Algorithms
  - Join-based algorithms
    - One spatial join per candidate colocation
  - Join-less algorithms

- Spatio-temporal
  - Which events co-occur in space and time?
    - (bar-closing, minor offenses, drunk-driving citations)
  - Which types of objects move together?