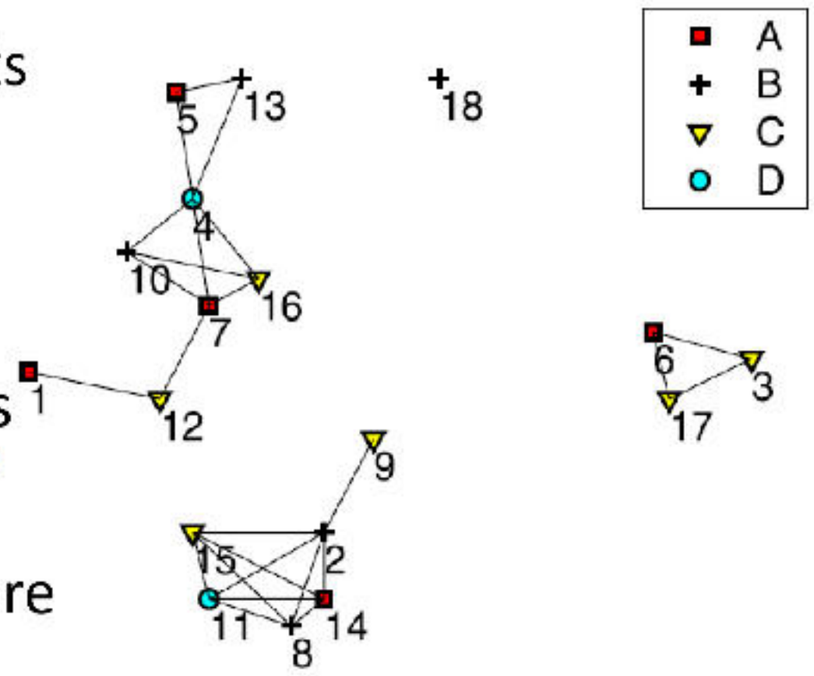
Mining Spatial Colocation Patterns

Spatial Colocation Patterns

- *Colocation pattern:* A group of spatial features or events that are frequently co-located in the same region
  - Ex. West Nile Virus often occur in regions with poor mosquito control and the presence of birds
- Figure: Neighborhood instances are connected by edges
  - Ex. {3, 6, 17}, {4, 7, 10, 16}, {2, 8, 11, 14, 15}, {2, 9}, …
- Rowset(C) if every feature in patter C appears as a feature of an instance in the neighbor-set L, e.g.,
  - rowset({A, B, C, D}) = {{4, 7, 10, 16}, {2, 11, 14, 15}, {8, 11, 14, 15}}
  - rowset({A, B}) = {{5, 13}, {7, 10}, {2, 14}, {8, 14}}
- A colocation rule R: A→ B, conditional probability cp(R) is defined as

$$cp(R) = \frac{|\{L \in rowset(A) | \exists L' \text{ s.t. } (L \subseteq L') \wedge (L' \in rowset(A \cup B))\}|}{|rowset(A)|}$$

  - cp({A, B} → {C, D}) = |rowset({A, B, C, D})|/|rowset({A,B})| = ¾ = 75%

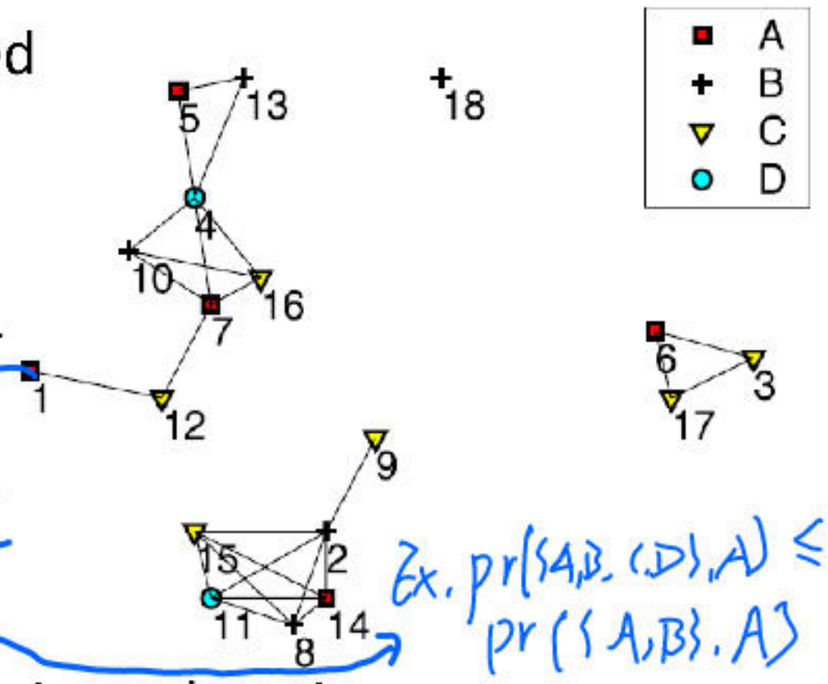


Mining Spatial Colocation Patterns

- *Participation ratio* *pr(C, f): probability that* C is observed in a neighbor-set wherever feature *f* is observed

$$pr(C, f) = \frac{|\{r|(r \in S) \wedge (r.f = f) \wedge (r \text{ is in a row instance of } C)\}|}{|\{r|(r \in S) \wedge (r.f = f)\}|}$$

  - Ex. *pr*({A,B,C,D}, A) = 2/5, ..., *pr*({A,B,C,D}, D) = 2/2 = 1
  - Monotonicity of *participation ratio*
    - Let *C, C'* be two co-location patterns such that *C' ⊂ C*
      - Then, for each feature *f ∈ C'*, *pr(C', f) ≥ pr(C, f)*



- An Apriori-like algorithm can be derived for efficient mining colocation patterns
  - Ex: Let min-feature-support = σ, min-*pr* = ρ
    - Start with a set of single feature pattern $\{p_1\}$ with support ≥ σ
      - Grow to size *k,* in Apriori way (i.e., stop growing if the pattern is infrequent)
    - For each such p, mine its super-pattern P, s.t., pr(P, p) ≥ ρ, in Apriori way