k-anonymity in go

Richard Garai (garric@okki.hu)

Anonymization in Business Systems - BSc
Consulent: Akos Dudas
BME AUT

April, 2019

privacy is ever more important

- tremendous growth in collection and analysis of personal data[1]
  - cloud services
  - mobile apps
  - smart fridge
  - etc. . .
- new regulations
  - EU: GDPR[2]

data anonymization

medical data

name	age	race	gender	zip	disease
John	47	white	male	1077	cancer
Sandy	35	white	female	1077	flu
Mary	27	asian	female	1095	flu
Janet	27	white	female	1095	hypertension

- identifier
- quasi-identifier
- non-identifier

anonymized medical data

name	age	race	gender	zip	disease
*	30..50	white	*	1077	cancer
*	30..50	white	*	1077	flu
*	27	*	female	1095	flu
*	27	*	female	1095	hypertension

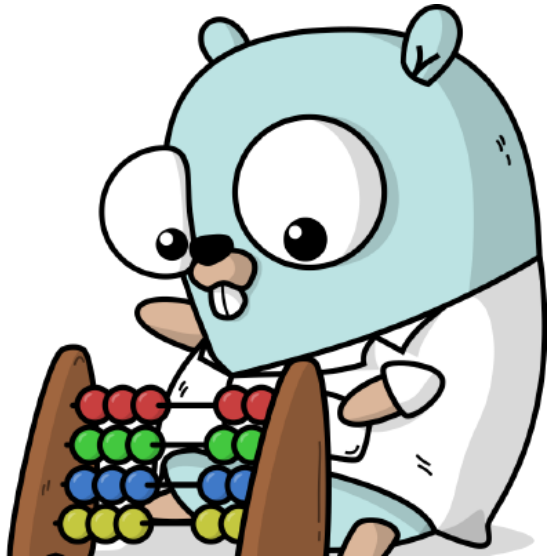
- suppression
- generalization

definition of k-anonimity

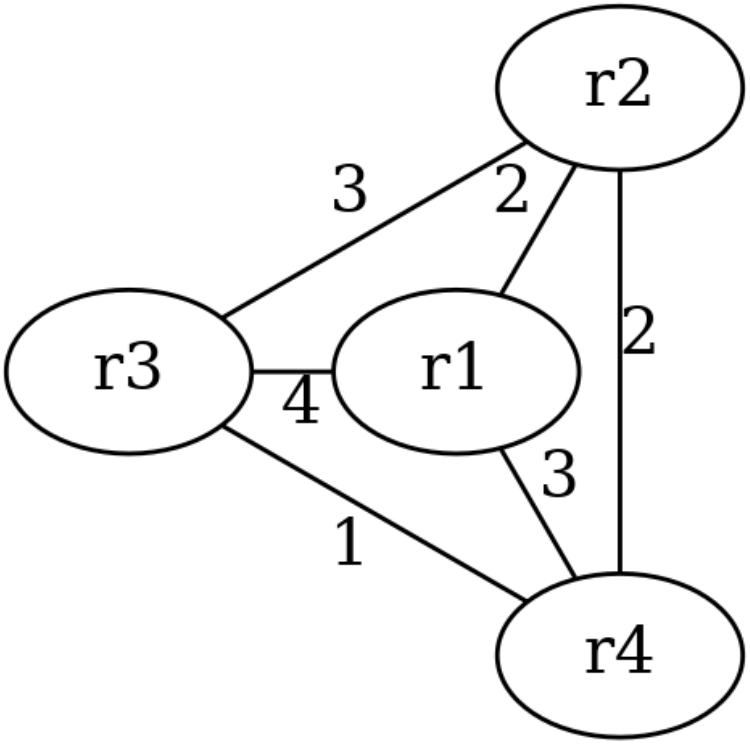
k-anonimity

suppress/generalize entries in the table until for each row, there are at least k-1 other rows that are identical to it along the quasi-identifying attributes

- k-anonymity even with only suppression and a ternary alphabet, i.e. $\Sigma = \{0, 1, 2\}$ is NP-hard [1]



cost graph

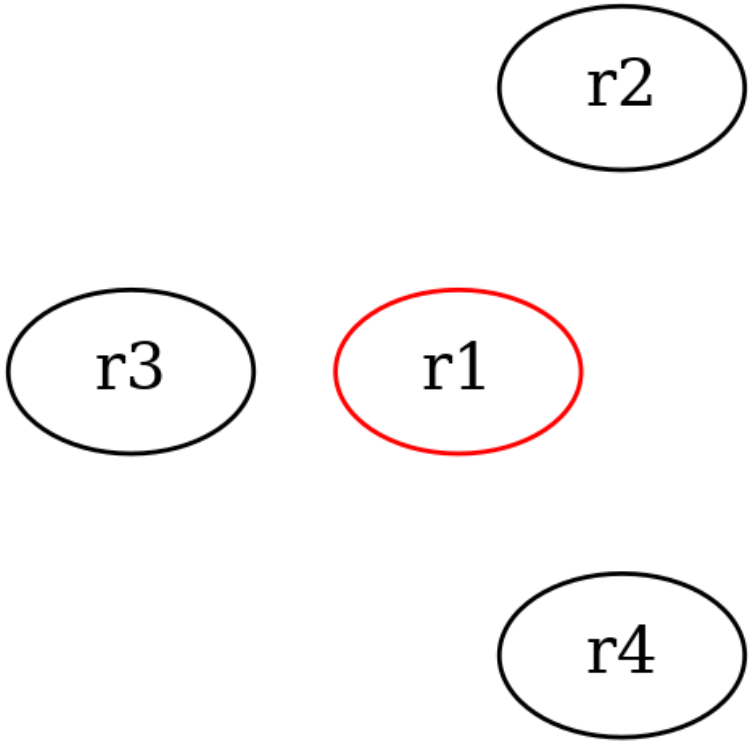


edge weight

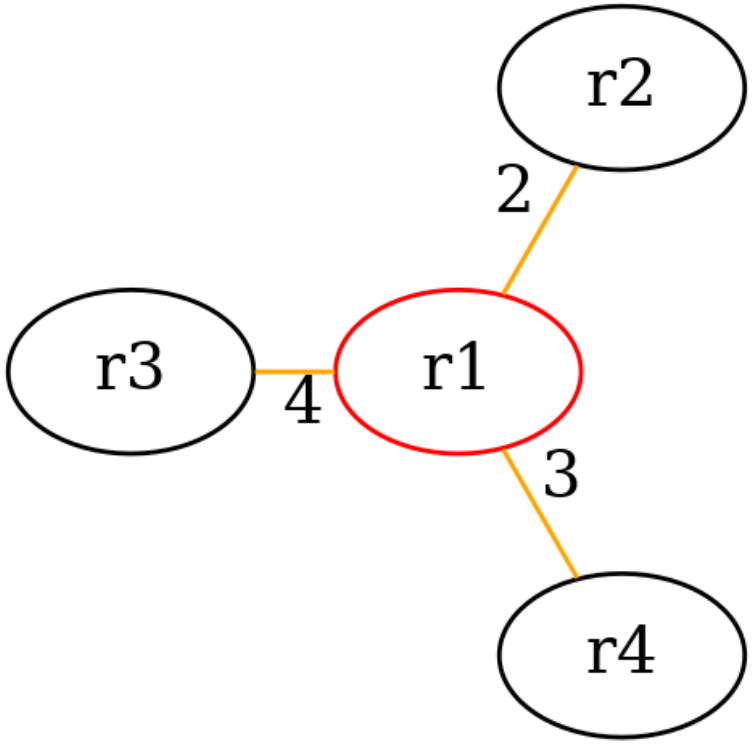
$$w(e) = \sum_j \frac{h_{a,b}(j)}{l_j}$$

$h_{a,b}$ generalization cost of items a and b
l_j maximum levels of generalization for column j

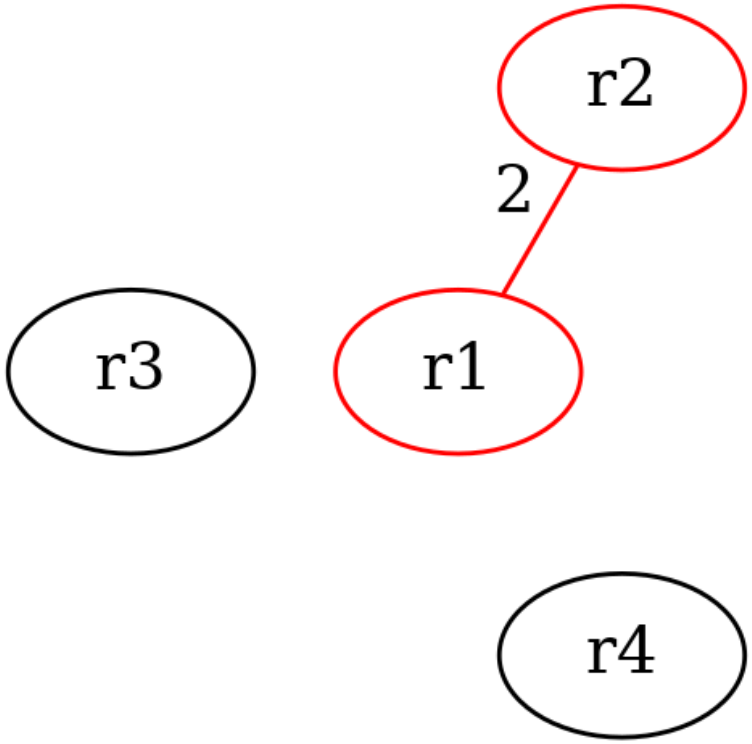
forest building



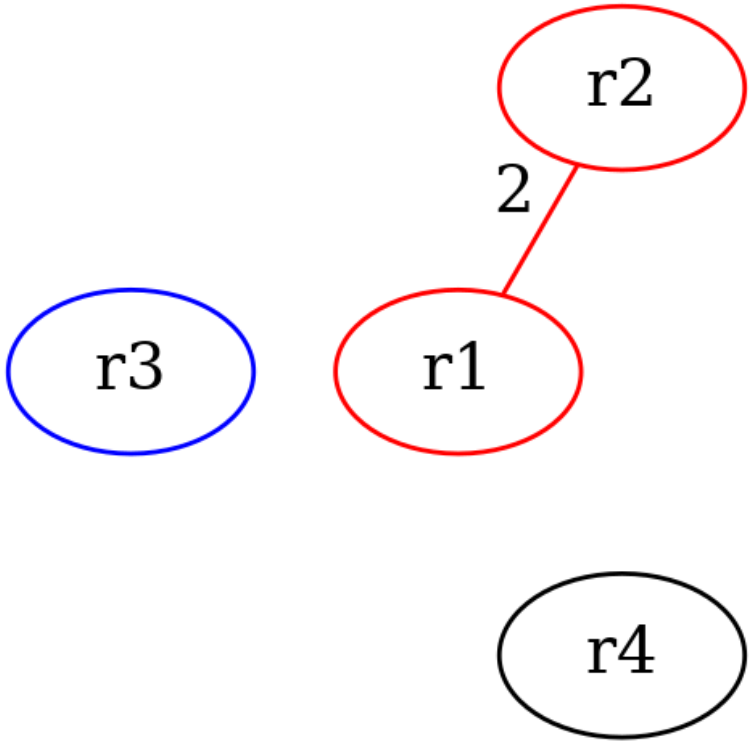
forest building



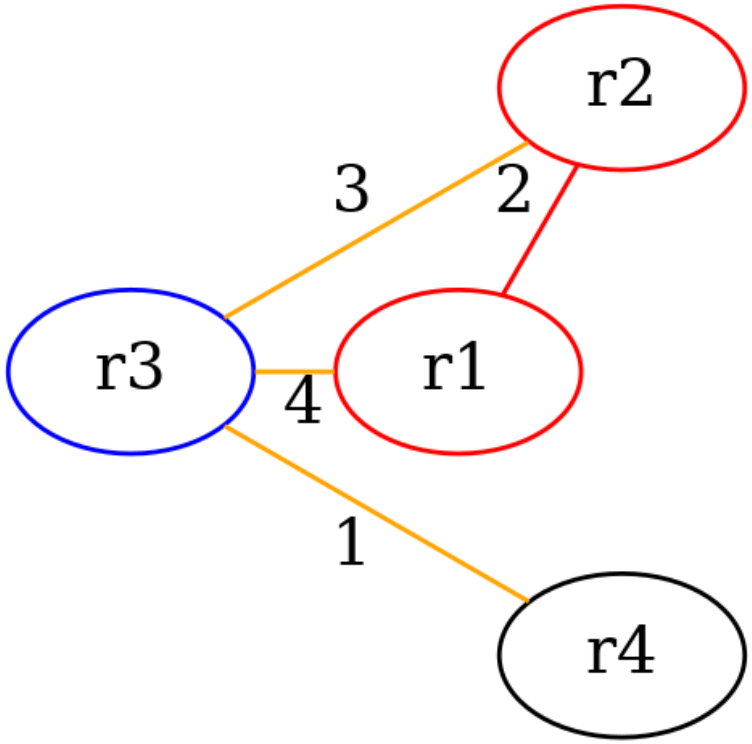
forest building



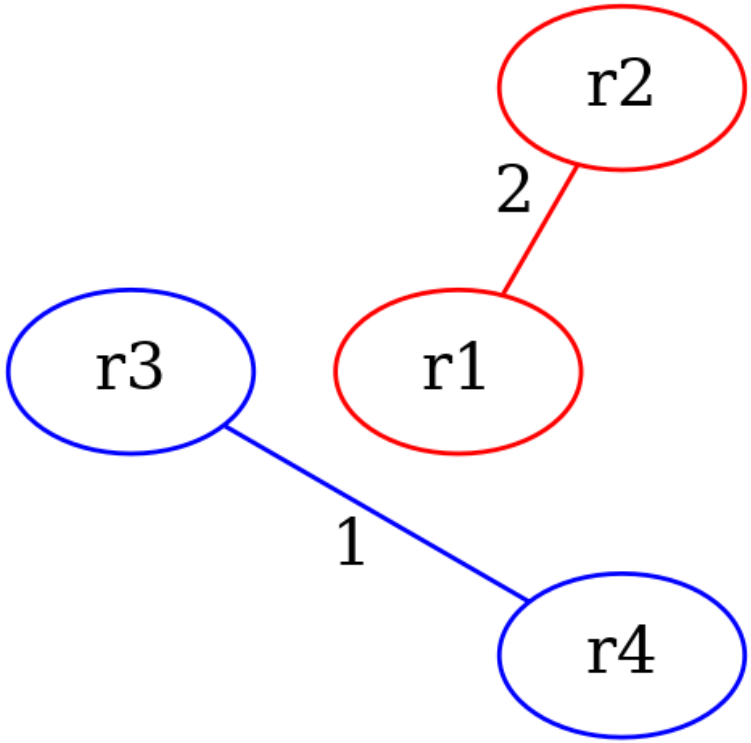
forest building



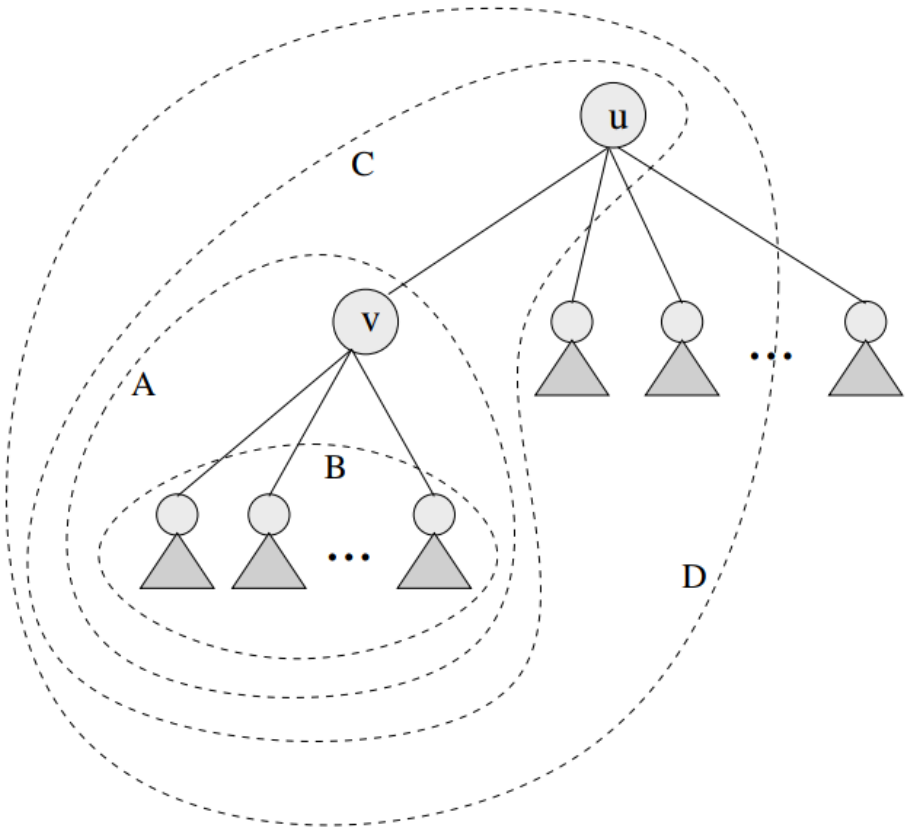
forest building



forest building



partition refinement



large components are refined further

$$s > max\{2k - 1, 3k - 5\}$$

- 4 distinct cut types
- Steiner's Vertices [1]

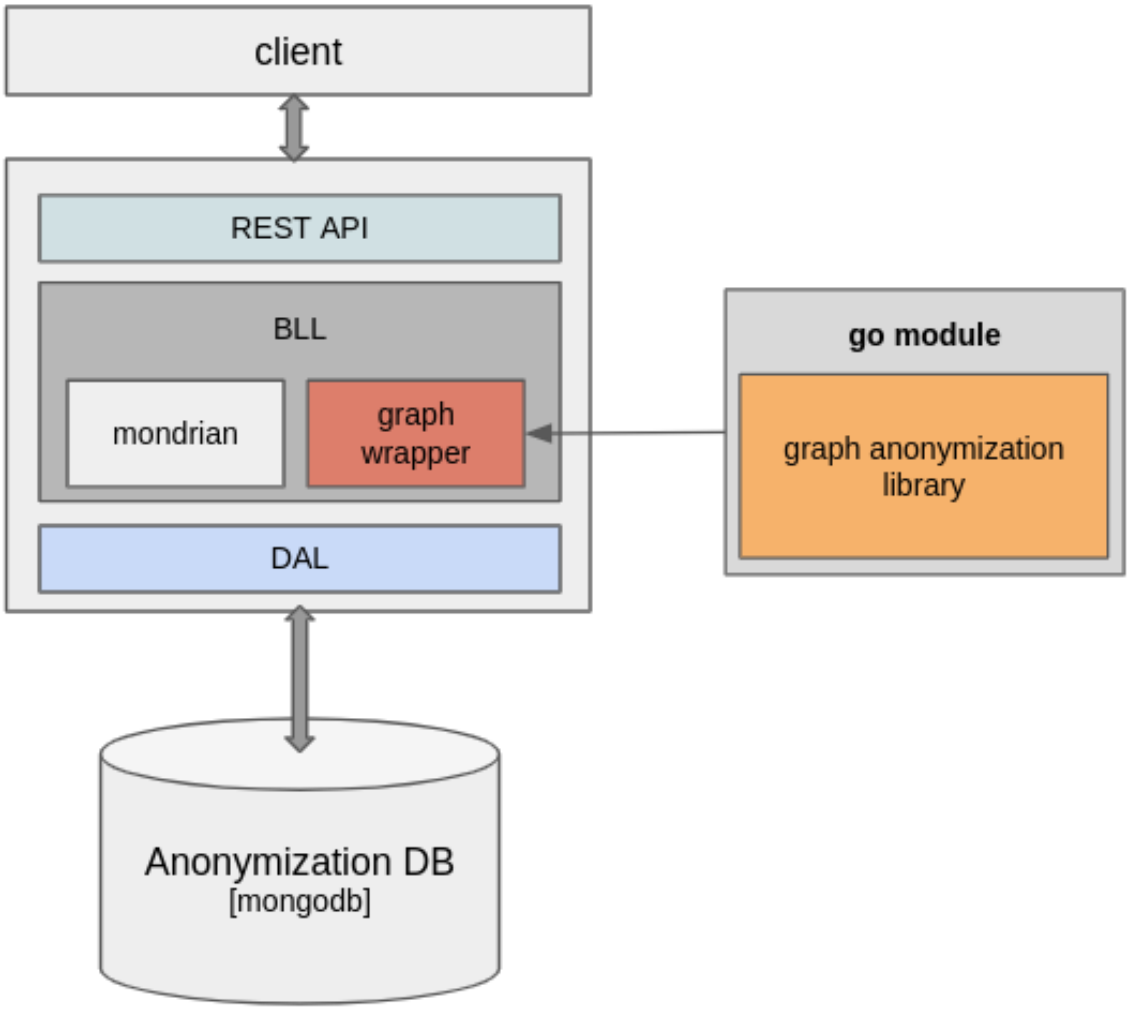
anonymization example code

```
table := model.NewTable(&model.Schema{
   Columns: []*model.Column{
      model.NewColumn("col", &generalizer),
      [...]
   },
})

table.AddRow(item1, item2, ...)


anon := &Anonymizer{
   Table: table,
   K:     2,
}
err := anon.Anonymize()
```

generalizers:

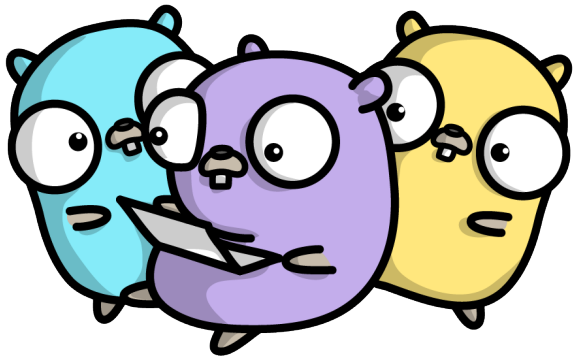
- suppressor
- int range
- float range
- string prefix
- hierarchy

integration with existing anonymization system



future improvements, references

- optimized algorithm for special cases
  - k = 2
  - k = 3
- continous anonymization (✓done)
- anonymization of streaming data



Approximation Algorithms for k-Anonymity
Gagan Aggarwal, Tomas Feder, Krishnaram Kenthapadi, Rajeev Motwani, Rina Panigrahy, Dilys Thomas, An Zhu Journal of Privacy Technology, 2015.

Data Anonymization
Wikipedia Article (https://en.wikipedia.org/wiki/Data_anonymization)