Advanced Behavioral Analysis Using Inferred Social Networks - Vision

Irena Holubová, Martin Svoboda, Tomáš Skopal,
David Bernhauer, **Ladislav Peška**

Faculty of Mathematics and Physics, Charles University, Prague, CZ

TIR 2019, Linz, Austria

Context & Disclaimer

- Industry-based project
  - Co-operation with a modest-sized bank & Profinit EU (data analytics)
  - General motivation: „get to know“ our clients better
    - Increased revenues, detect potential problems, unwanted situations etc.
  - Rather a fuzzy task without strict goals
  - *Work-in-progress (more questions than answers ☺)*

- Traditional banking
  - Relational data (clients, products, transactions)
  - Standard analytics (SELECT/GROUP BY/FILTER), ML based on facts
  - Some statistical task-specific analytics
    - Mostly predictions in risk management (loan/mortgage business), fraud detection/prevention
    - Common ML techniques, simple features (demographics, risk classes, income, etc.)

Motivation

- What can be done beyond traditional approaches?
  - Go beyond explicit fact checking
  - Model similarities of bank clients based on their behavior
  - Construct some latent „social network“ based on clients' similarities
    - Social relations, locality relations, financial behavior patterns, etc.
    - No explicitly given task -> focus on new business cases
      - Personalized marketing, insurance recommendation, investments, financial consultation, insolvency prevention, etc.

GOALS (bank's point of view):

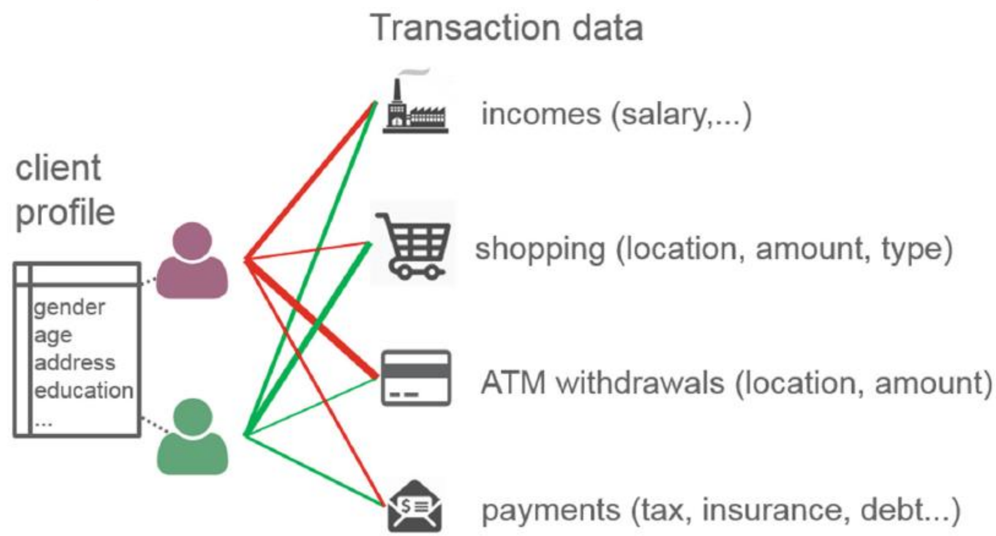
- *Offer appropriate financial products to relevant clients* ***at the right time***
- ***Predict & prevent unwanted situations*** *(frauds, loan payment problems)*
- ***Find new business models***

Motivation

- Kate just added travel insurance to her card
  - William is similar to Kate w.r.t. expenses structure, withdrawn ATM’s countries and location patterns.
    - *Human readable explanation: they both travel but don’t like extreme sports => basic insurence will do. Also, maybe they’re a family?*
  - *What about recommending travel insurance to William as well?*
  - ***Further confirmation:** in past, recommendations along these axes worked well.*
- Miloš did not received his salary last two months
  - His expenses remain similar & he keeps withdrawing from „booze ATMs“ cluster
    - *Human explanation: he lost his job, but he keeps drinking like hell*
  - *What about offering him a loan?*
  - **Maybe, not a good idea:**
    - Miloš’s expenses structure & withdrawn ATM’s clusters are similar to how did Andrej behaved a year ago. Andrej received a loan, but failed to repay it.

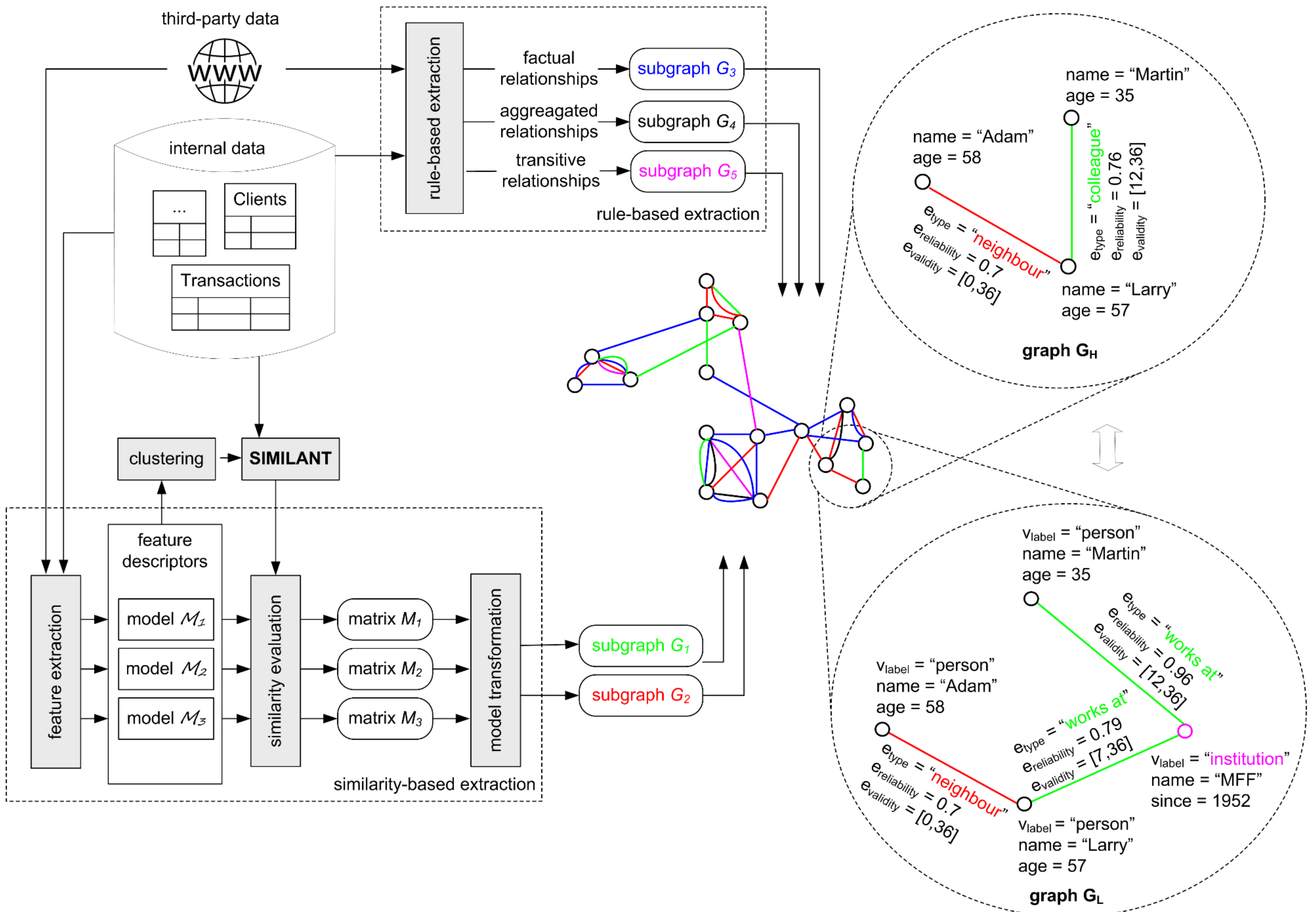
Data available

- Demographics
  - Name, gender, age, address…
  - Credit score, loan payment history, frauds…
  - *Some data anonymized / hashed for us*
- Transactions
  - Allow to model similarities of clients based on various financial behavior profiles
  - Rich attribute structure
  - Temporal context

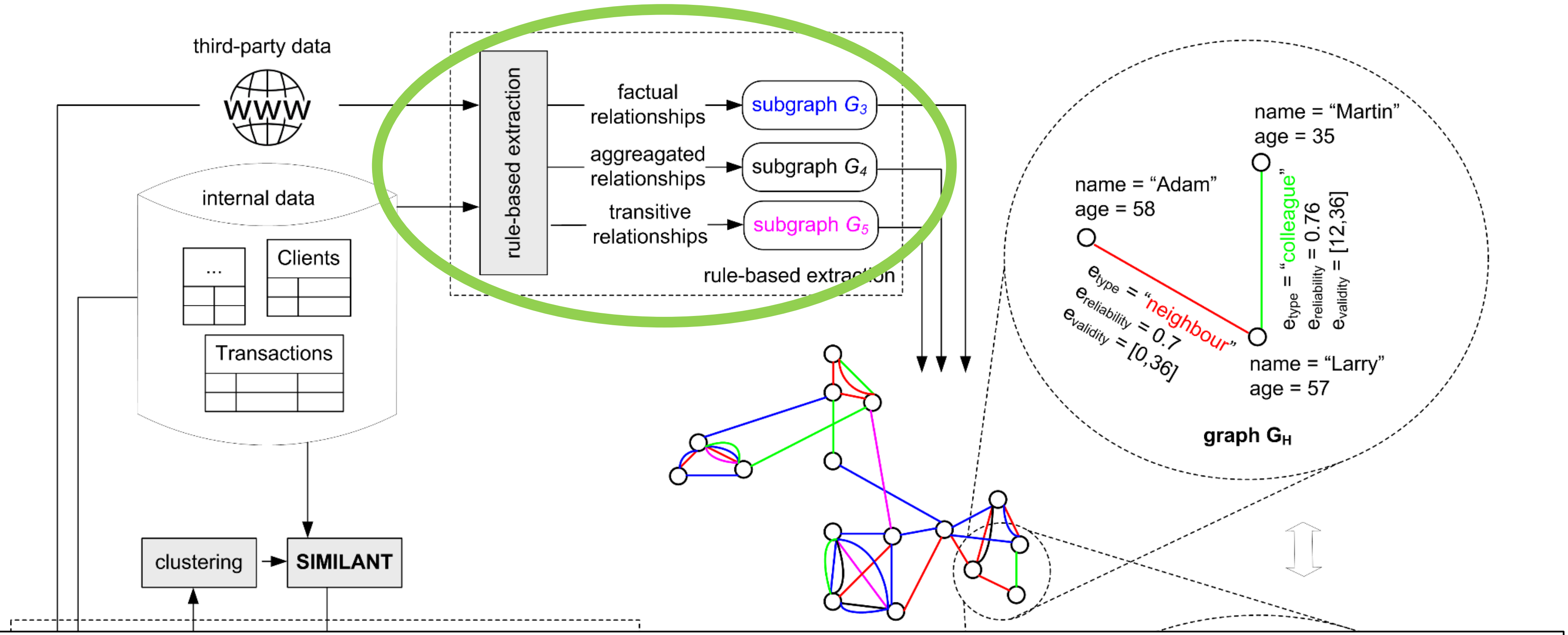


- 3rd Party Data
  - E.g. Real estate, business or criminal registers
  - Legal concerns (however useful for manual verification)

Latent social network model



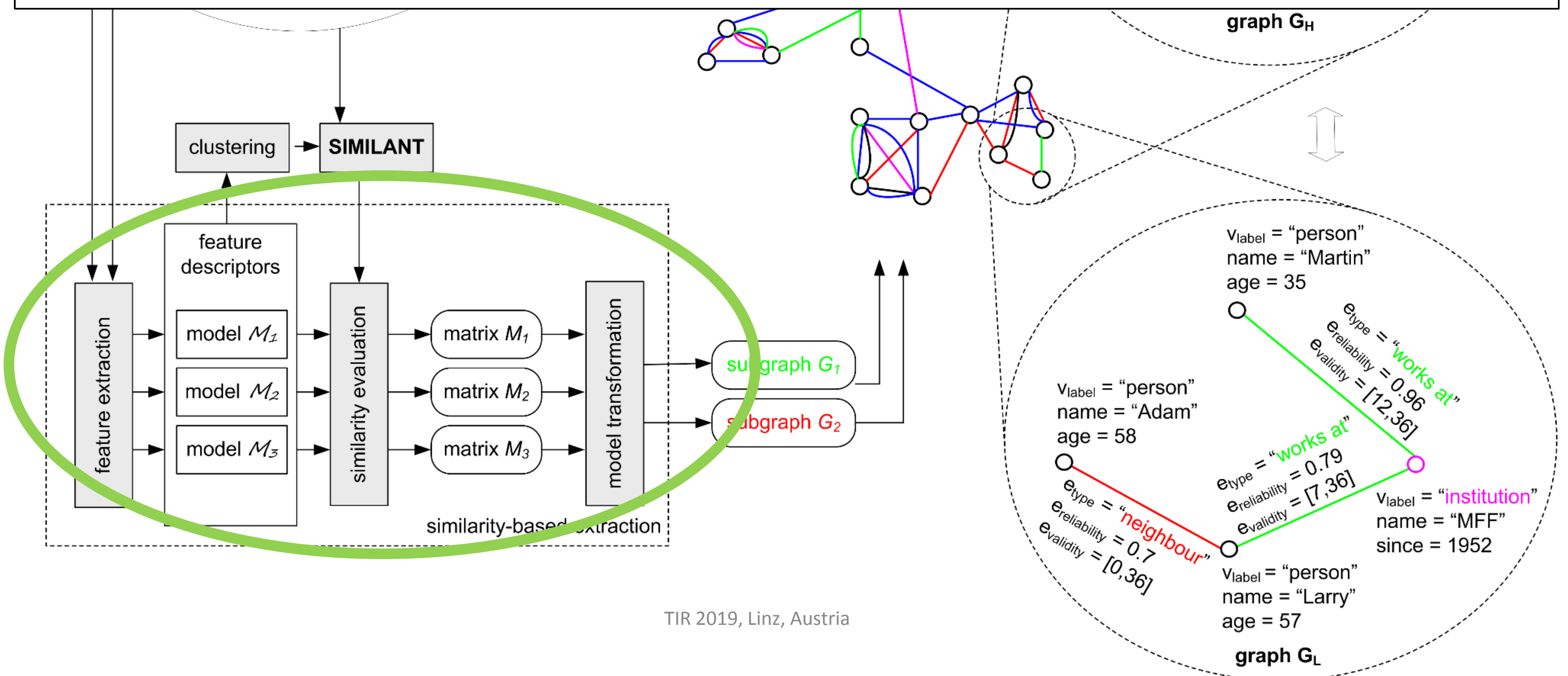
Latent social network model



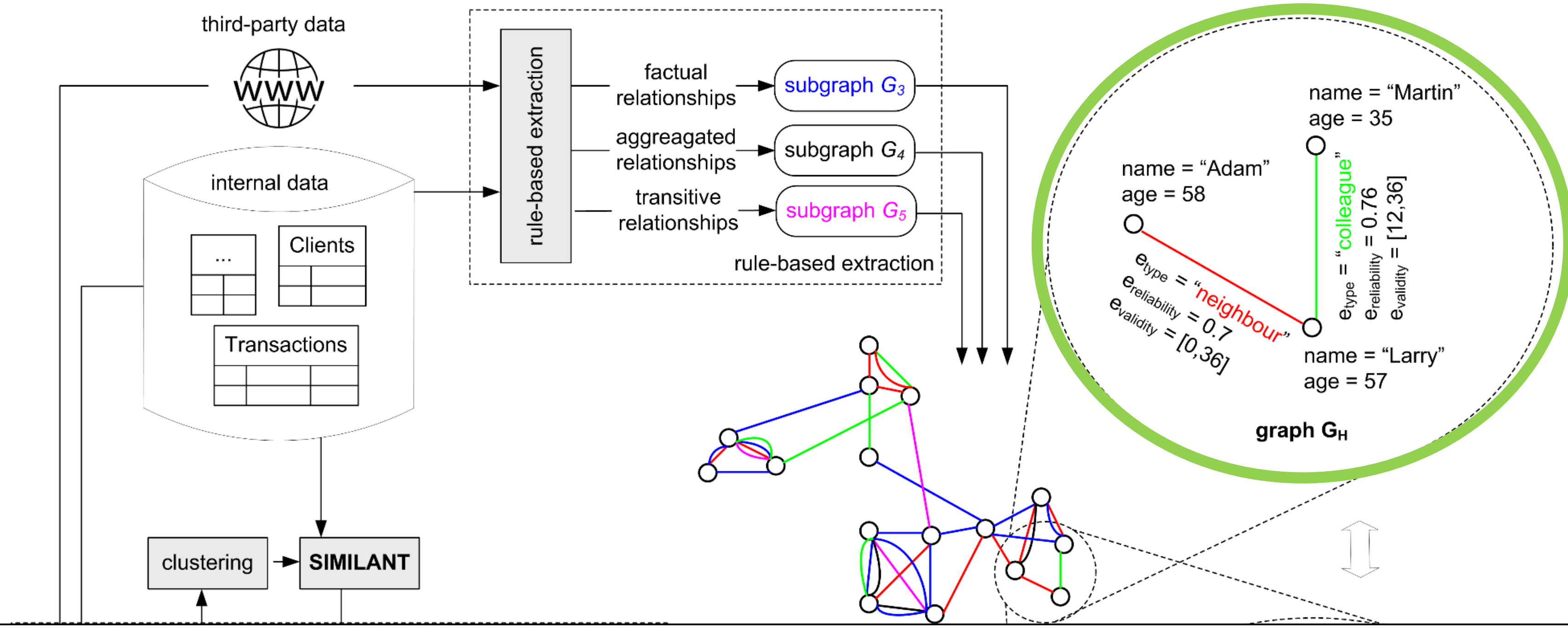
- Rule-based edge extractions
  - Fact checking, e.g., family/relatives, co-workers
  - Aggregated relations (e.g., mutual transactions -> business relations / employers)
  - ***Can be done with traditional approaches***

Latent social network model

- Similarity-based edge extractions
  - Focus on client's behavior
  - Stream of transactions (possibly window-based aggregations)
  - Amount, datetime, location, frequency, counterparty, transaction type, type of business (+ brand) etc.



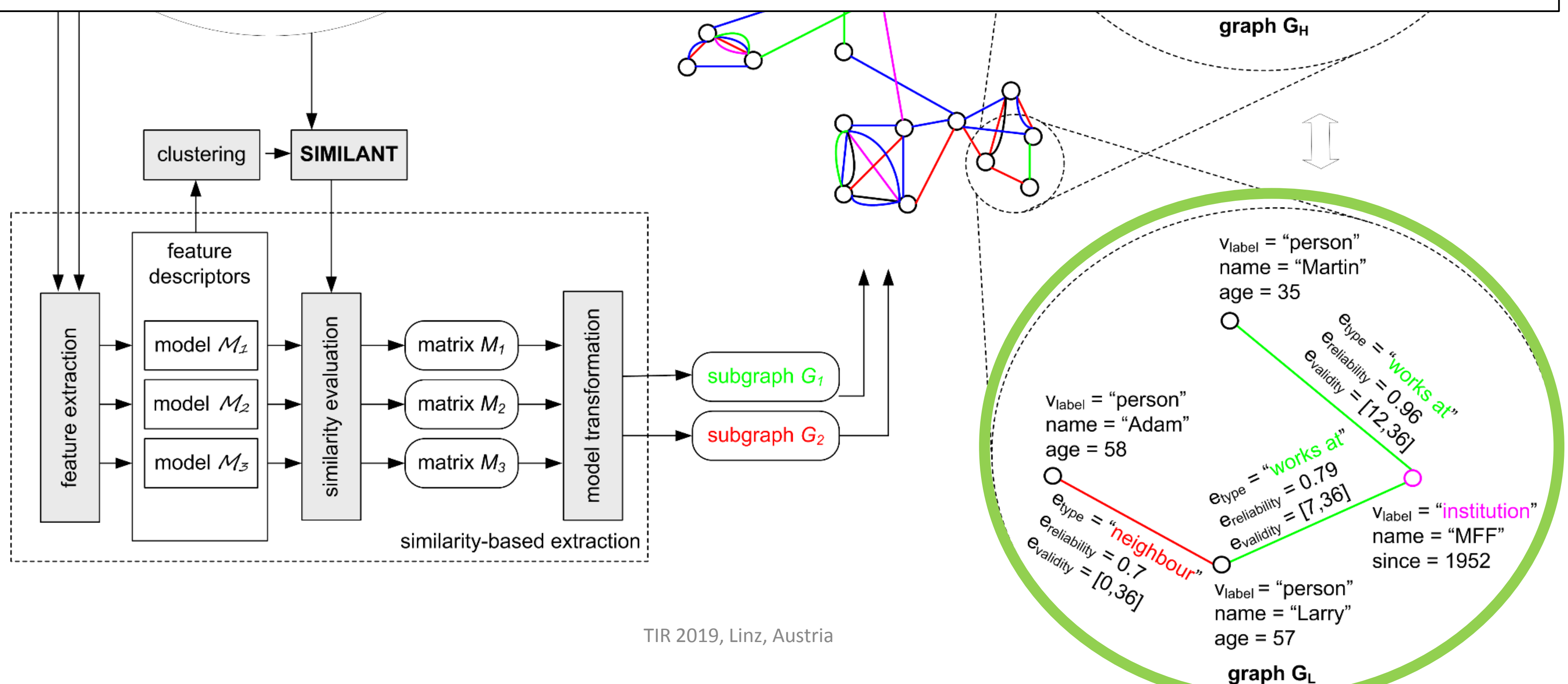
Latent social network model



- High-level latent social multigraph
  - Relations among clients (aggregated from low-level edges)
  - Client’s attributes (both factual and inferred)
  - Time-aware edges, relation relevance score

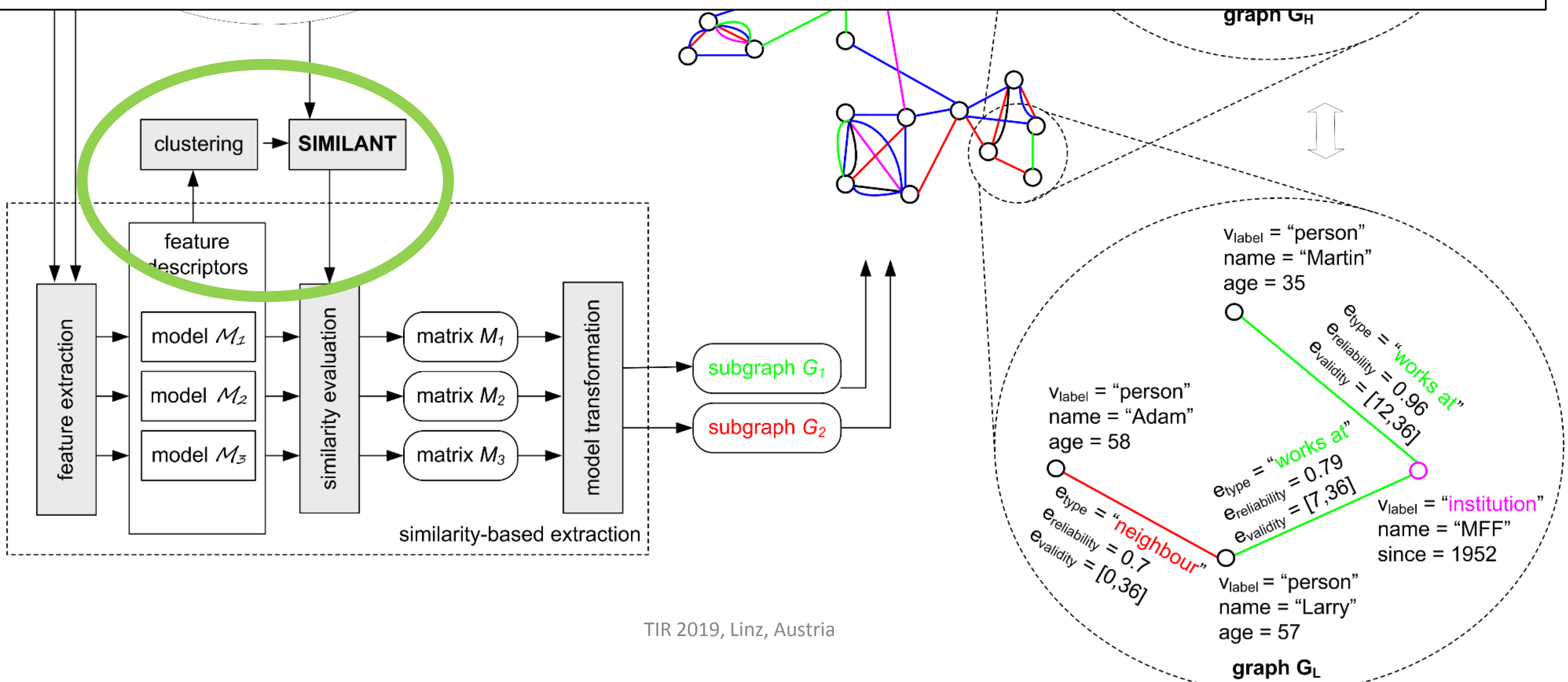
Latent social network model

- Low-level latent social multigraph
  - Decomposed relations over other node types (locations, institutions etc.)
    - E.g. Person1-Colleague-Person2 ->
      Person1 – works_at – Institution & Person2 – works_at - Institution
  - Finer grained relationship mining

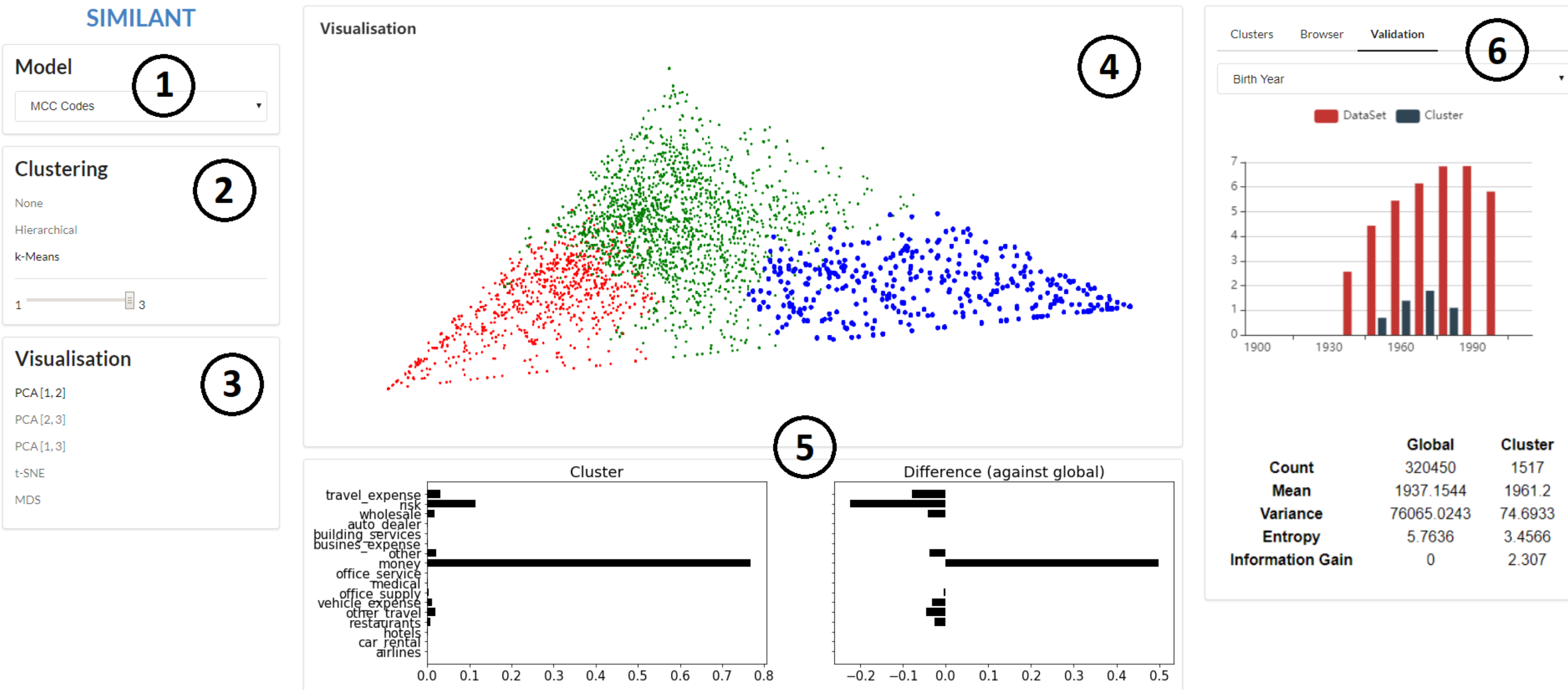


Latent social network model

- SIMILANT
  - Analytic tool to evaluate similarity descriptors
    - *Which descriptors are meaningful? What pre/post-processing? What similarity metric?*
    - Clustering & visual evaluation, inferred node's attributes
    - Possible automated validation w.r.t. available targets & e.g. Information gain

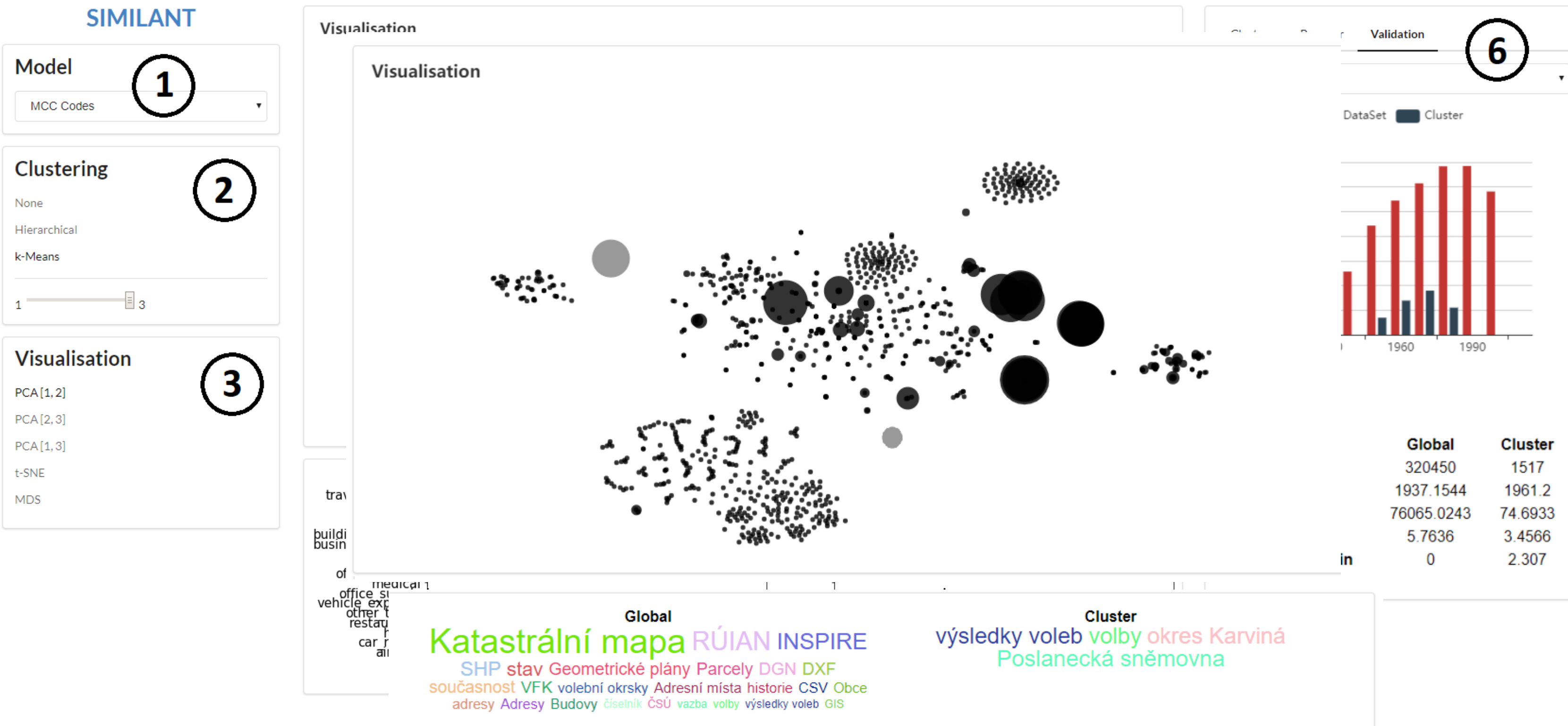


SIMILANT



- 1-3) Select descriptor&similarity metric, clustering, visualization
- 4-5) Visualize clusters and features of the descriptor
- 6) Browse selected cluster, its instances & validate over known targets

SIMILANT



- Several visualization variants
  - Various per-cluster features can be mapped to cluster size and color
- Several options to display & compare cluster descriptors

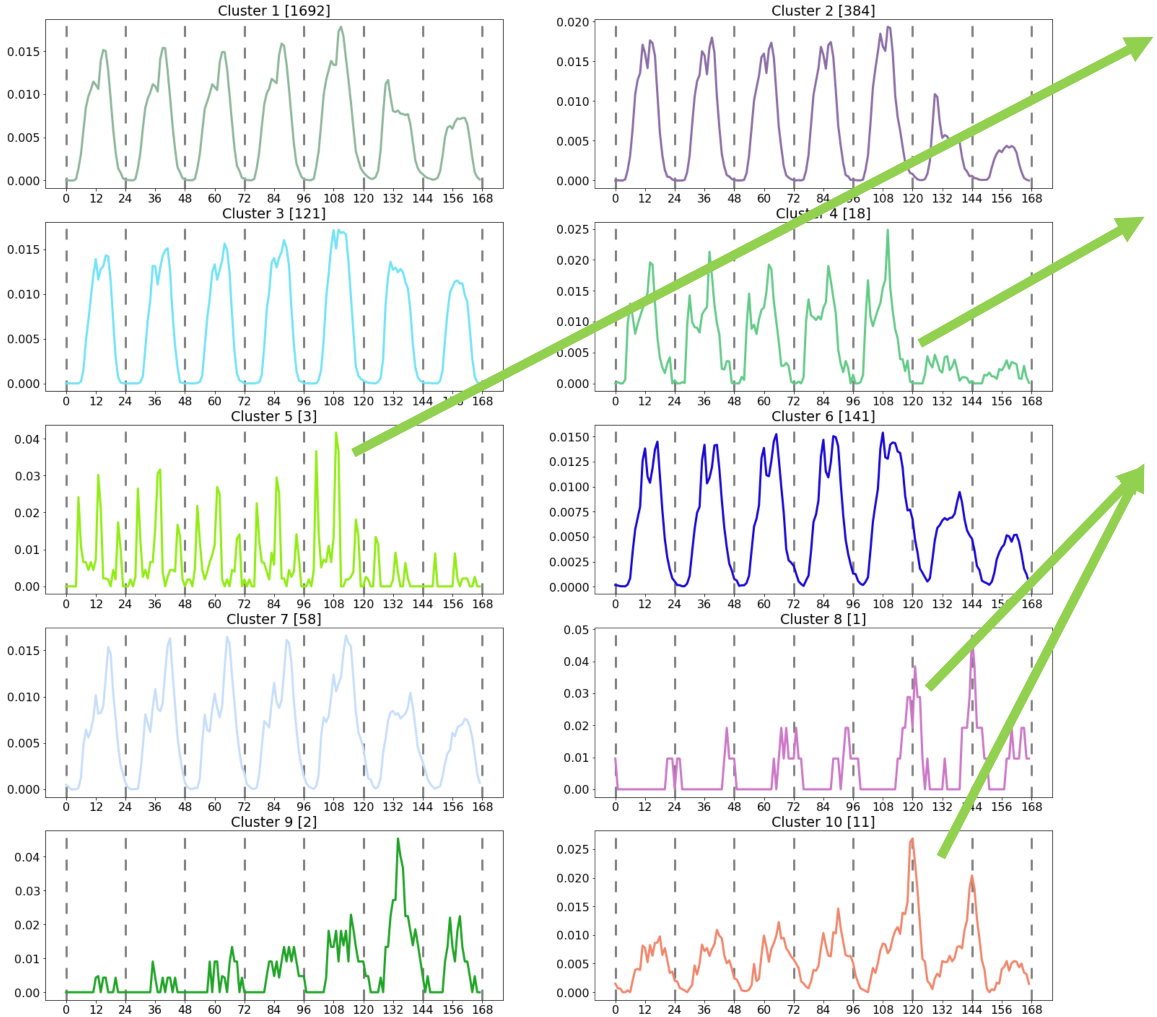
Do we have relevant knowledge?

- Heavy anonymization caused several problems
  - Lack of proper targets (what to validate against?)
- Final solution may differ from some current proposals
  - household detection might be just a hidden attribute
- Dataset is quite noisy (e.g., outdated addresses)

- *Nonetheless, still capable to disclose interesting features*

Some examples so far

ATM's withdrawn amount weekly per hour



3-shift factory workers

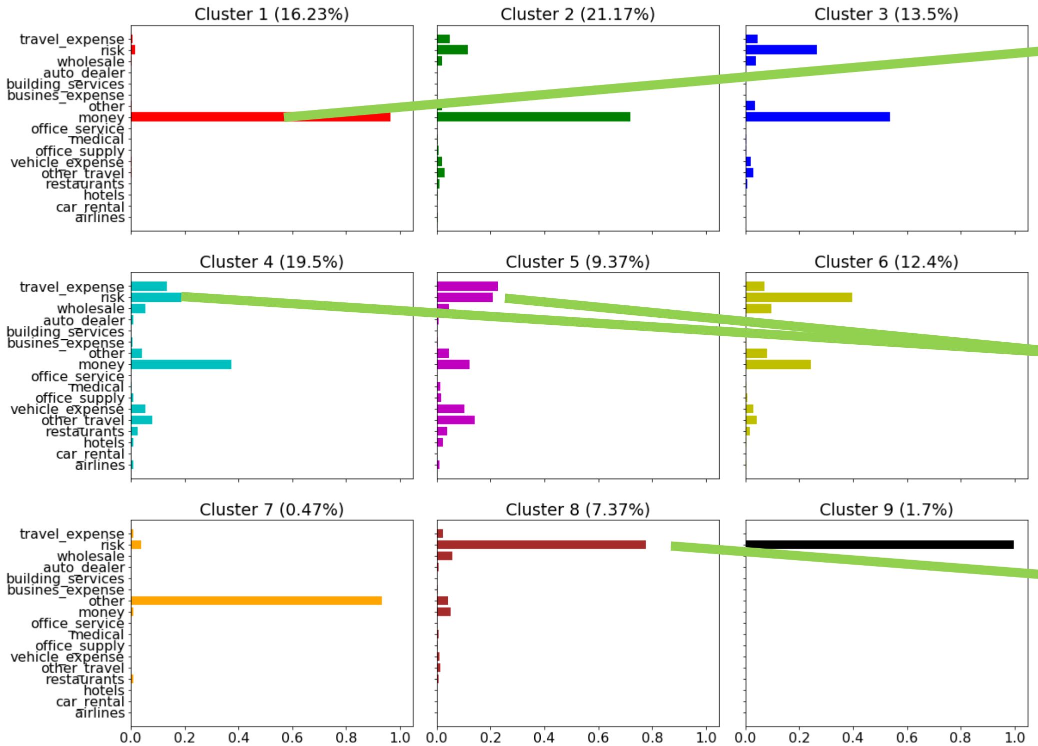
Business centres (drop on weekends)

„booze ATMs“ (close to bars)

Another example: high average night withdrawals for an ATM close to a night club

Some examples so far

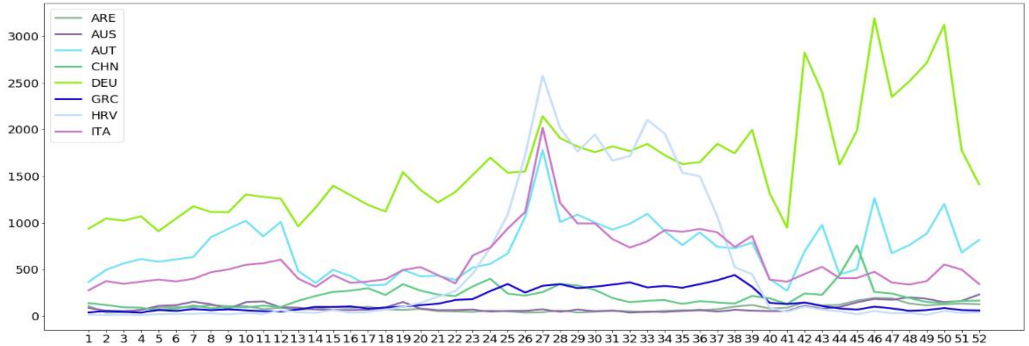
Card payment MCC codes



Mostly ATM withdrawals – do not like card payments / elderly persons?

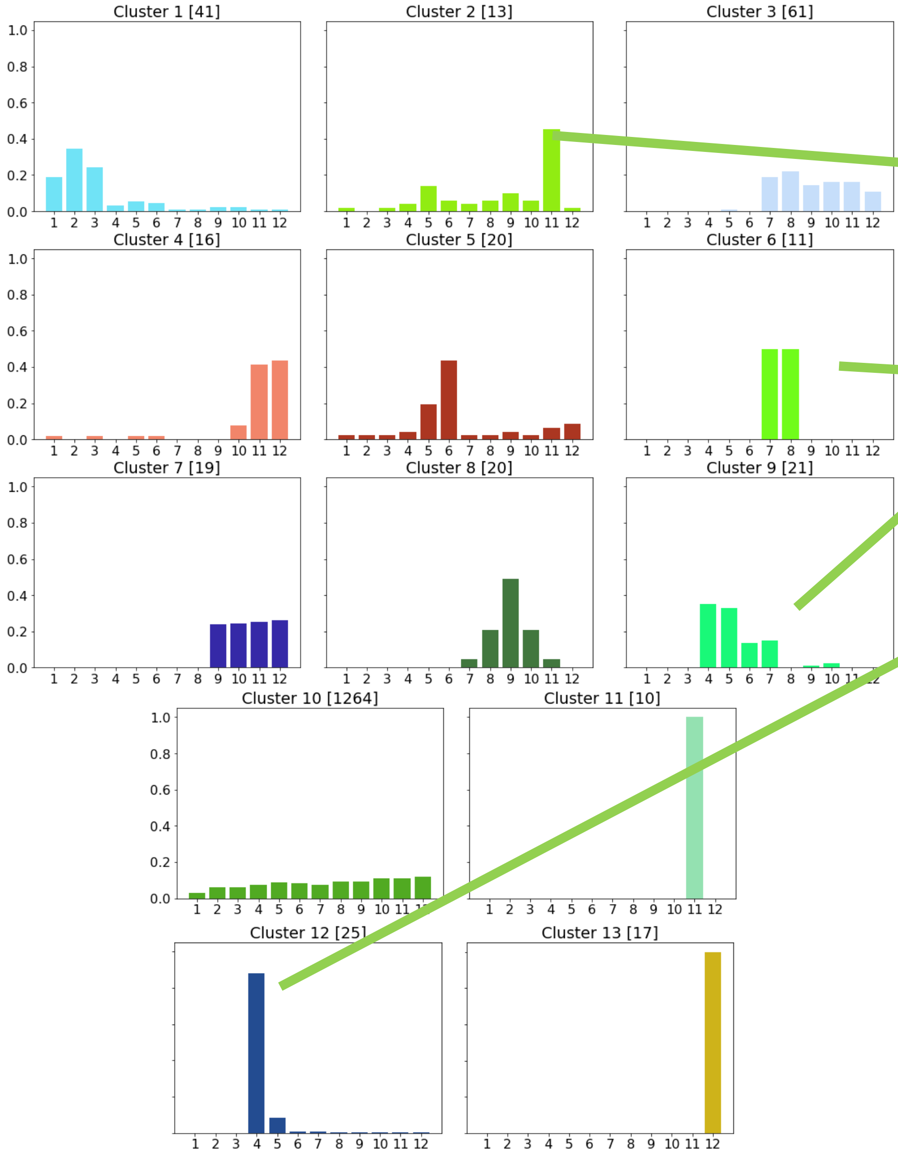
Frequent travellers

Paying by card for common things



Weekly payments in foreign countries (holiday peaks): distinguish between travels for work and leasure

Some examples so far



Counterparties: per month outcomes

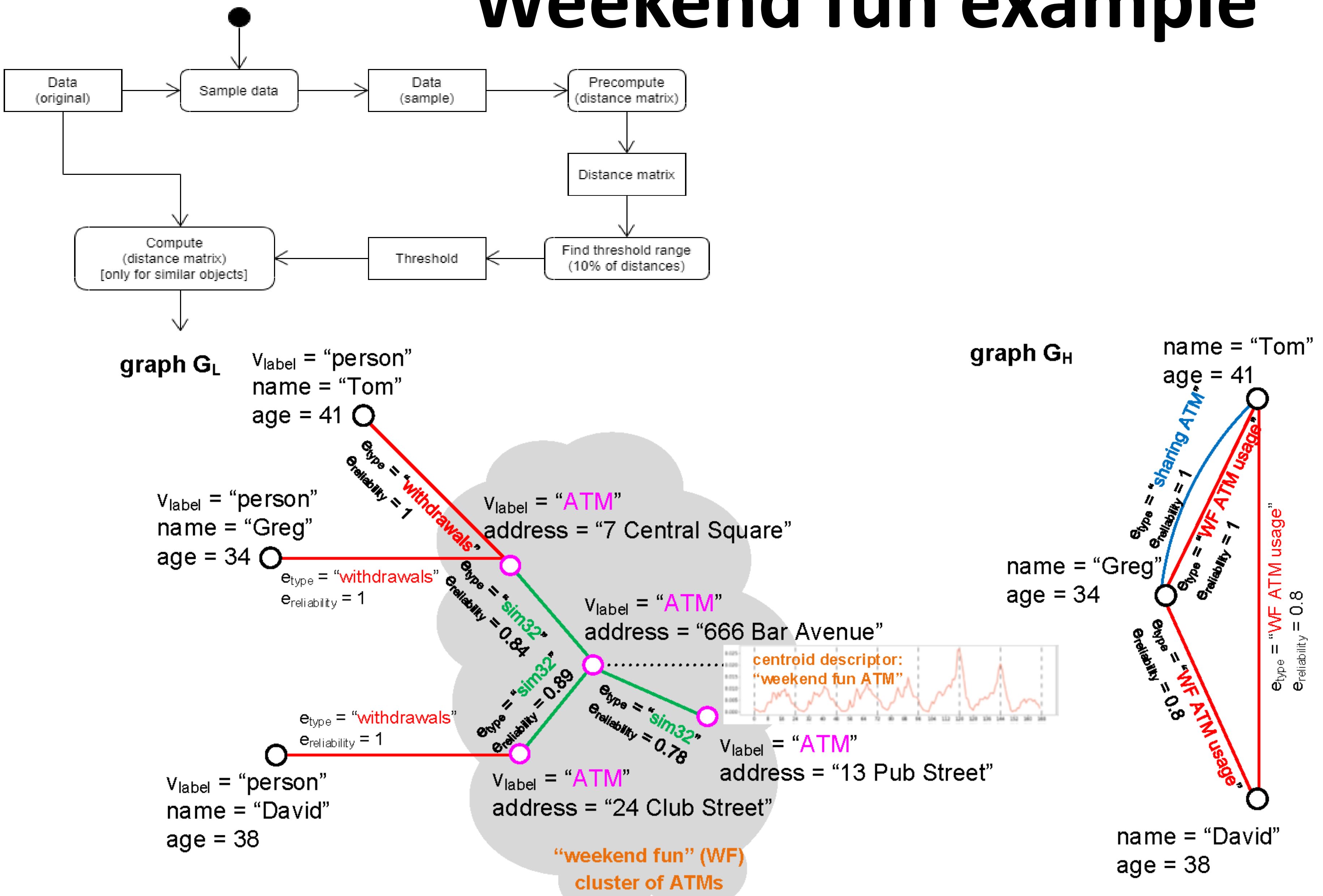
Christmass loans?

Seasonal labor?

Financial office (tax refounds)

No need to explain everything, just to believe that similarity is meaningful

Weekend fun example



„sim32" = DTW on hourly withdrawals

Current state & Future work

- Preliminary network construction process ready
- SIMILANT tool for evaluating individual similarity descriptors
- Some early results seems promissing
  - Partially explored: client features, counterparties, countries, ATM withdrawals
  - TODO: merchant descriptions, payment patterns, locations etc.
- Challenges
  - A bit too broad domain (too many possible hypotheses, pre-processing, descriptors, similarity metrics, clustering & parameters)
  - Dynamic domain, detection of changes (life milestones, business closures), time-aware edges
  - **Ethical challenges** - reliability of latent edges, validity for important decisions
- Future plans
  - Compare with std. social network properties (communities, hubs?)
  - Time-aware models
  - Network visualization & exploration
  - **Explanations & action recommendations**
  - Expansion beyond banking domain (insurance, teleco etc.)

Thank you!
Questions?