Collection of Visualizations

EuroVis 2020 STAR

Supplementary Material

Objective

The objective of this document is to collect all the visualizations described in the selected corpus of papers, along with descriptions wherever required.

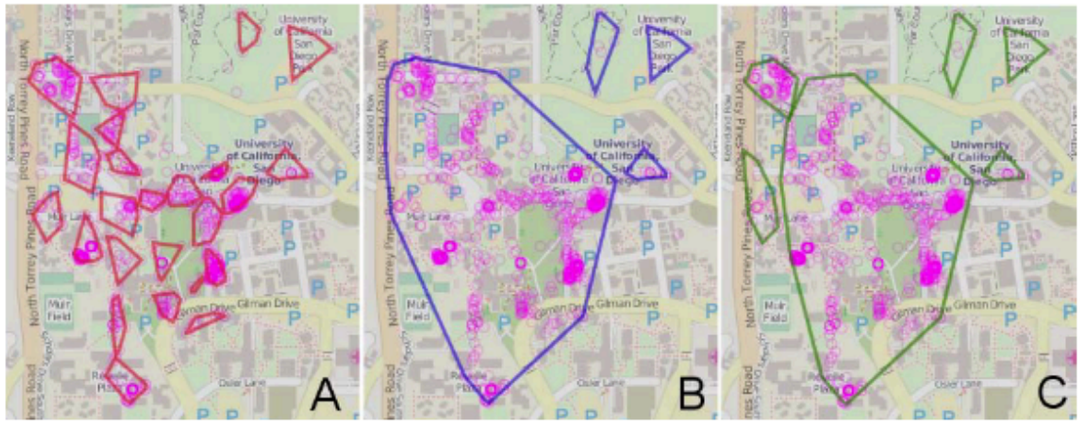


Fig. 5. Neighbouring point clusters may be united into larger clusters.

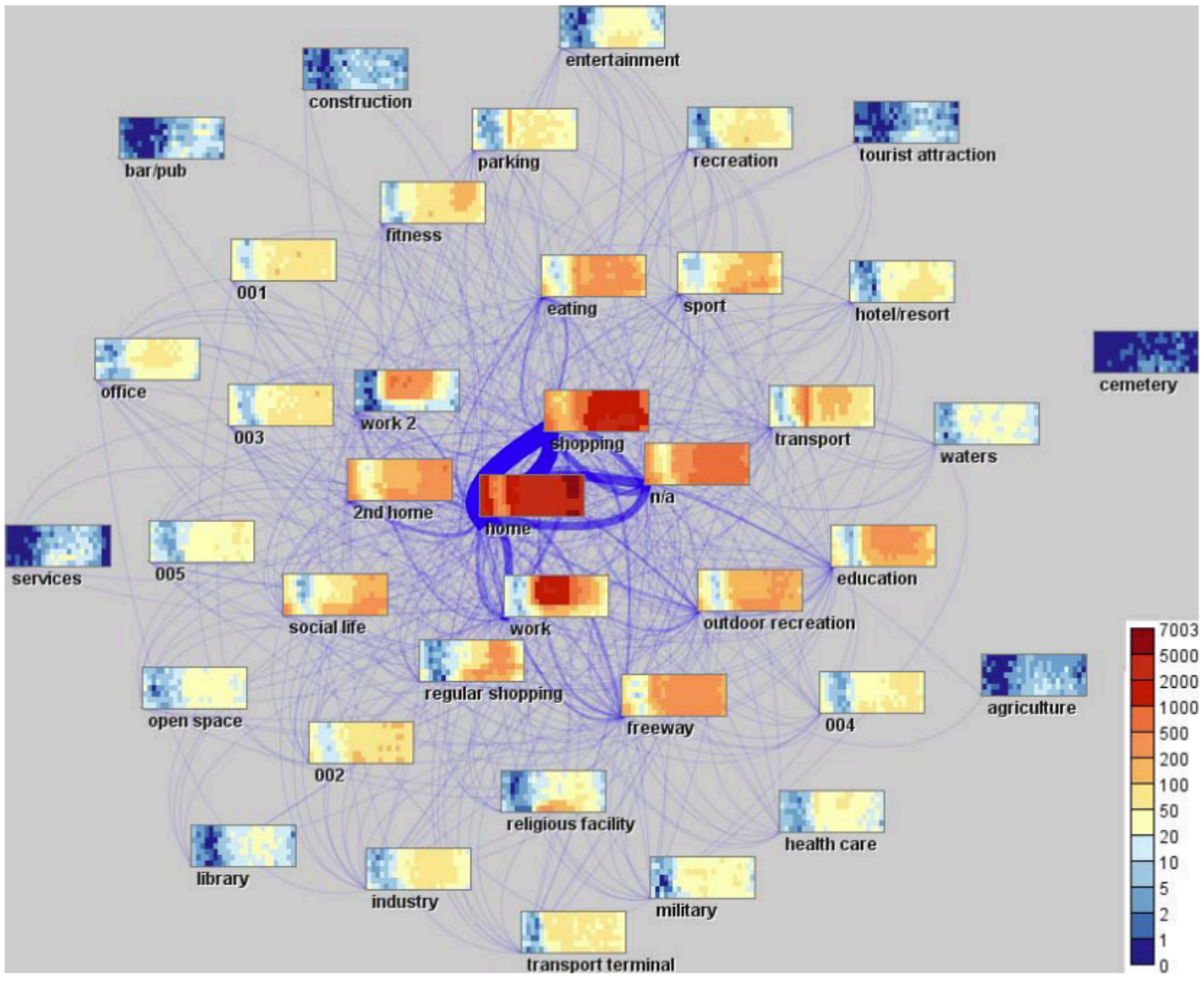


Fig. 17. Semantic information derived from the San Diego data is represented as a semantic space map.

[Andrienko2016] [AAFJ16] This paper presents a visual analytics model which can analyze the episodic digital traces/locations of a person over a long period of time and detect places of significant interest like home, work, social activity place etc. But this model also preserves the privacy of the person being analyzed. This paper uses geographical maps to cluster neighboring point clusters into larger. It also uses a semantic map to display the semantic information about different places derived from the data of a certain city. This paper also uses 2D time histograms to analyze the usage of different location clusters in a certain city over a certain period of time.

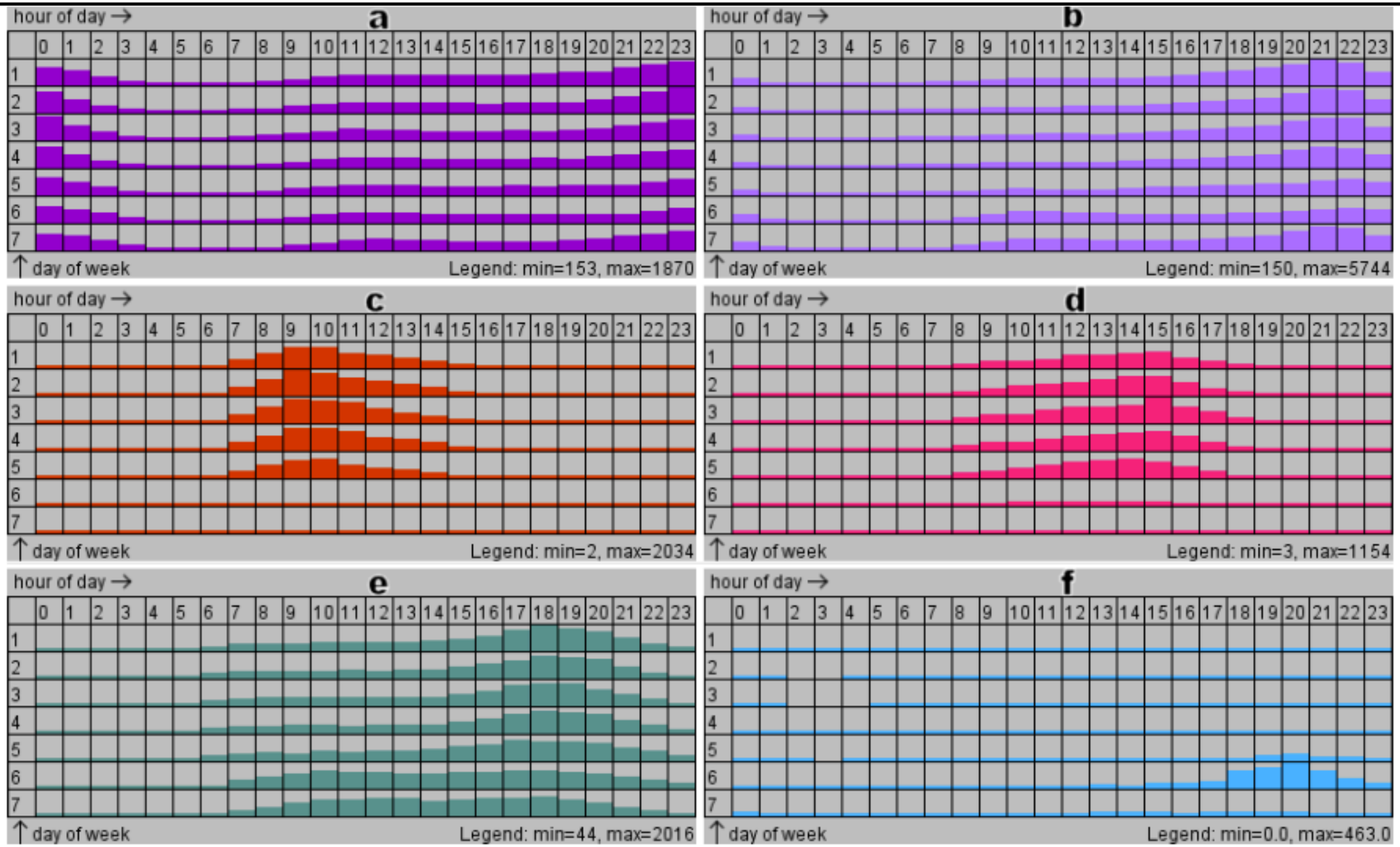


Fig. 10. 2d time histograms for selected clusters of places (San Diego example).

[Andrienko2016] [AAFJ16]

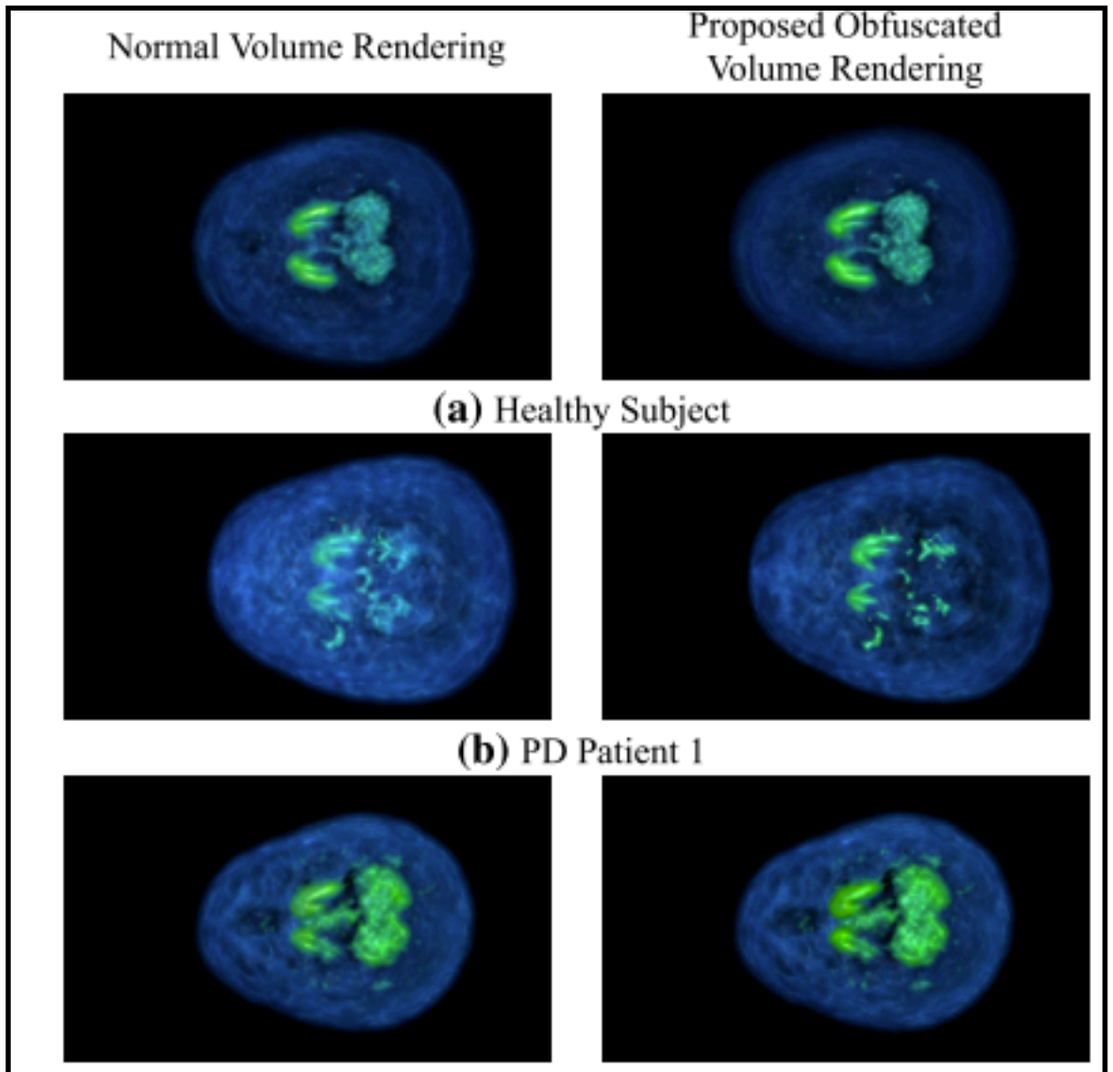


Fig. 12 Comparing the volume rendering results of human brains between a healthy subject (**a**) and two PD patients (**b**) and (**c**). In addition, each human brain is rendered by both normal direct volume rendering technique (shown on the *left* column) and the proposed obfuscated volume rendering approach (shown on the *right* column). It shows that the proposed method yields comparable rendering images with the normal rendering technique. As can be seen in **a** and **b**, it is obvious that the shape and silhouette of bilateral caudate, NAc, APu and PPu are more symmetric and visible in the health subject (**a**) than in PD Patient 1 (**b**). However, when a PD patient is with mild symptoms (PD Patient 2), the reduction of VMAT2 may not be obvious enough for one to tell whether the subject suffers from PD or not. In such a case, it may require us to extract a specific slice(s) of the data for further investigation

[Chou2016][CY16] This paper proposes an obfuscation technique for scientific visualizations in order to maintain the privacy of the user. This block based volume data transformation algorithm obfuscates a volume data and delegates the task of rendering the volume data to a remote server, thus preserving the privacy of the scientific visualization. The images show the difference between normal rendering and the proposed privacy-aware volume rendering. This paper has also developed a transfer function adjustment so that the transfer to the remote server for volume rendering is also privacy preserving.

(a) Showing the initial dataset of five people.

(b) The **Rookie** and **1+yrs** nodes ared merged together.

(c) The **Lab** and **Home** nodes are merged together.

(d) Edge bundling **[Person]** nodes connected to the **Lab** and **Home** nodes.

Figure 3: Privacy preserving operations: (a) The original dataset. (b) Applying a node merge to address k-anonymity. (c) Applying a node merge to address l-diversity. (d) Edge bundling to address l-diversity.

[Chou2017][CMB17] This paper focuses on creating privacy preserving visualization using ontological social network data. It suggests that methods like node and edge deletion, node merging and edge bundling in a node-link diagram can improve their privacy, this resolving k-anonymity and l-diversity privacy leaks. This paper also comments on the improvement of the readability of the ontology visualization in order to "allow viewers effectively perceive, analyze, and interpret an underlying dataset and it s salient features". This paper has also included two case studies demonstrating the effectiveness of these privacy preserving visualization techniques. The case studies were conducted on the MIT Reality Mining Dataset and the School Kids Friendship Dataset.

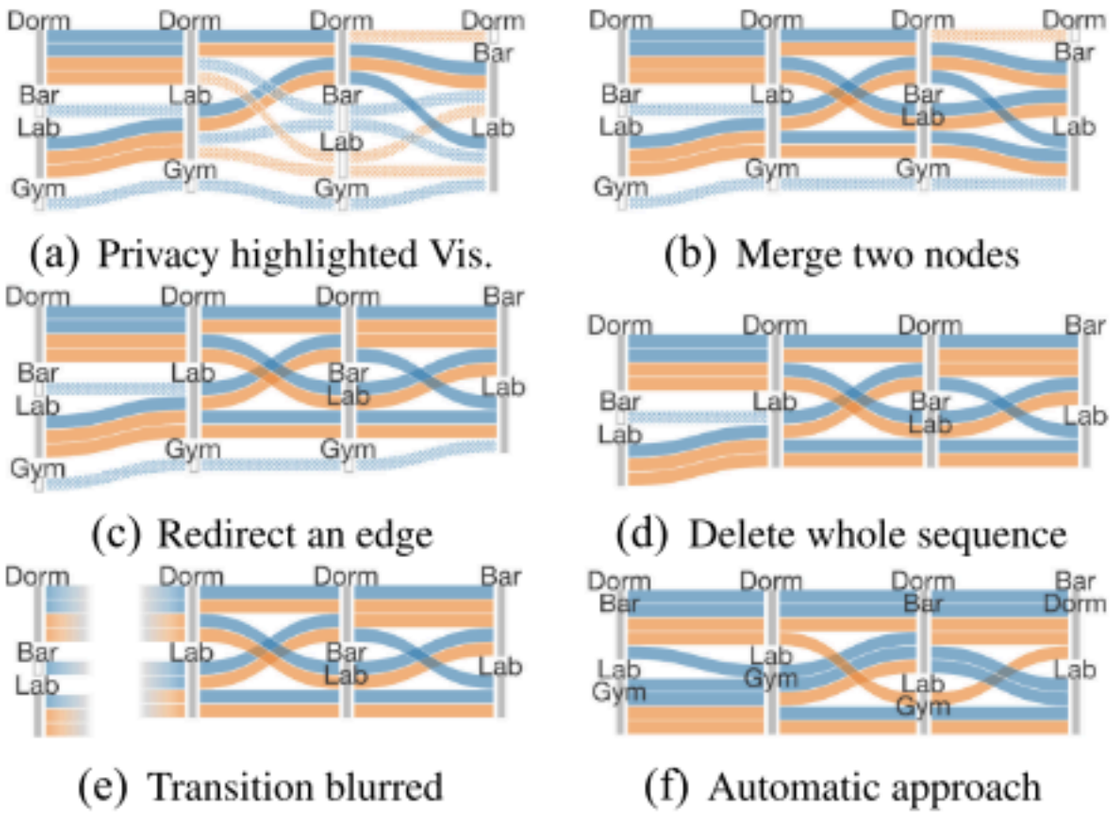


Figure 3: *Examples of applying the privacy preserving operations on the sample data set. (a) Highlighting the detected privacy issues. (b) Merging the "Bar" and "Lab" nodes at* t_2*. (c) Redirecting the destination of an edge from "Dorm" to "Bar." (d) Deleting the entire sequence of a subject. (e) Blurring the transition between the first two time points. (f) Addressing all the privacy issues highlighted in (a) by only considering the Merge operation.*

[Chou2019][CWM19] This paper focuses on developing privacy-preserving visualizations on event sequence data.It also discusses the privacy threats like identity disclosure and attribute disclosure and the privacy models like k-anonymity, l-diversity and t-closeness which help to overcome these privacy threats. This paper also discusses the privacy preserving operations on Sankey diagrams like merge, redirect, delete and blur, which will also help in retaining the utility of the visualization. It also offers the option of automated privacy preservation to users in order to resolve the privacy issues automatically throughout the whole visualization. Use cases based on MIT reality data, EMR data and a simulated daily schedule data have also been demonstrated in this paper.

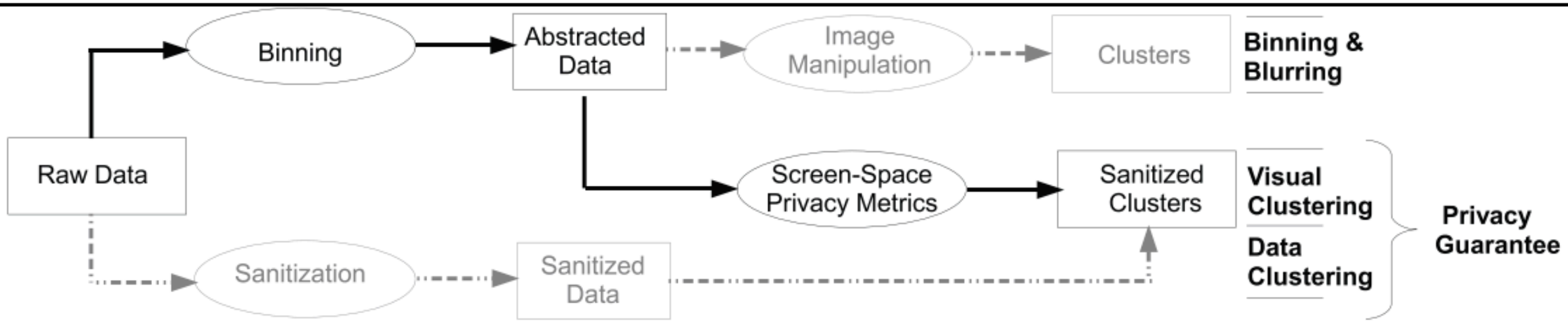
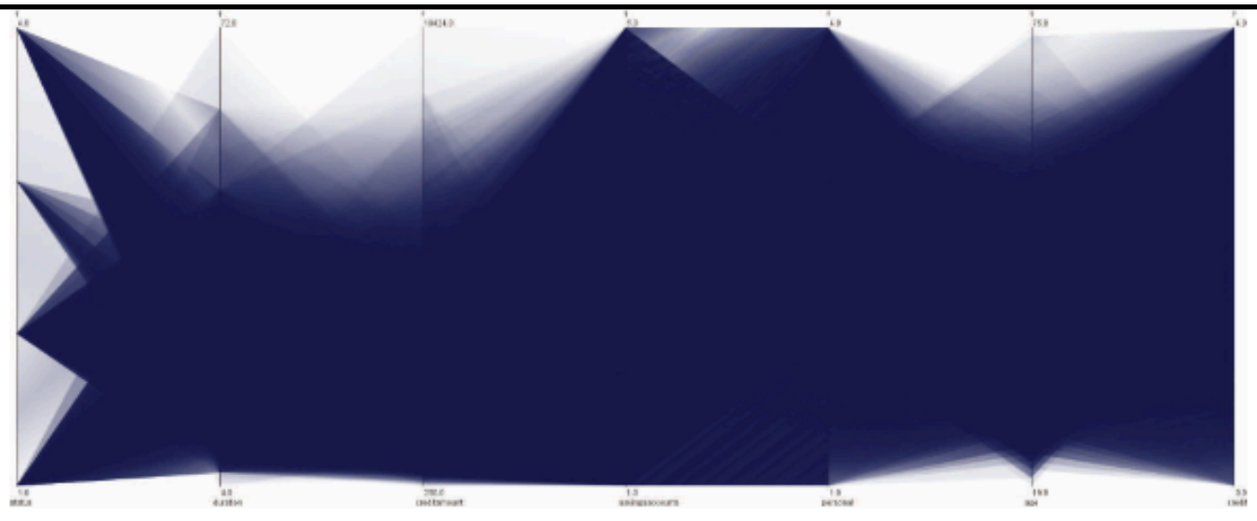
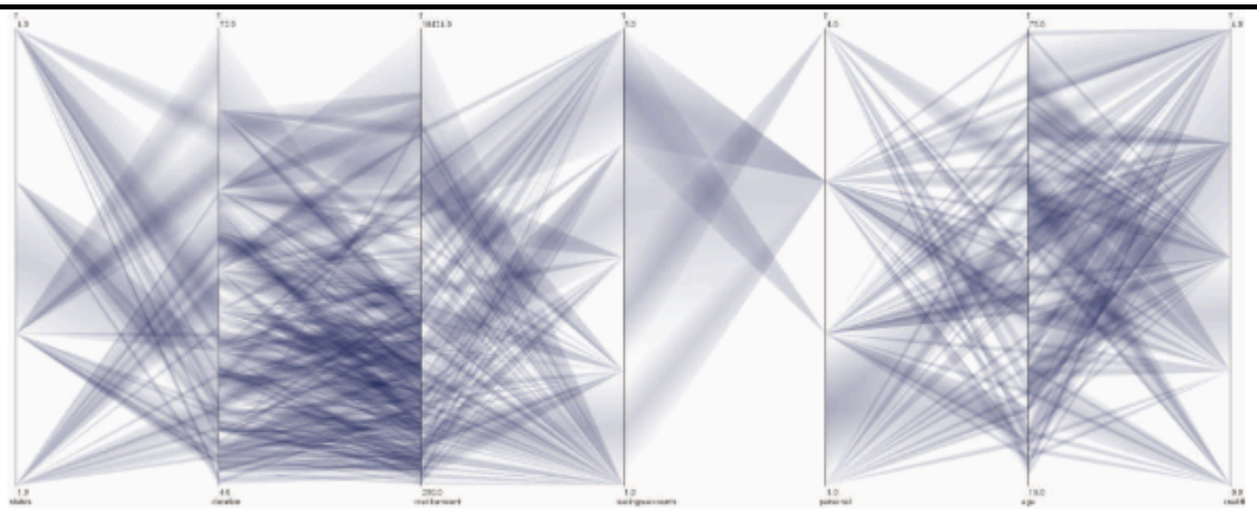


Fig. 1. Possible approaches to privacy-preservation in information visualization (Section 3.2): binning and blurring, data clustering, and visual clustering. Only the bottom two approaches guarantee a given level of privacy, in the case of data clustering, it leads to considerable loss of utility.



(a) Visualizing data that was sanitized using approaches from privacy-preserving data mining results in poor utility.



(b) Clustering by axis pair and using a different metric in the clustering, more of the visual structure can be retained.

Fig. 2. Comparing data clustering to our visual clustering approach. Both algorithms make it impossible to tell fewer than three records apart, but our approach provides higher utility.

[Dasgupta2011a][DK11a]

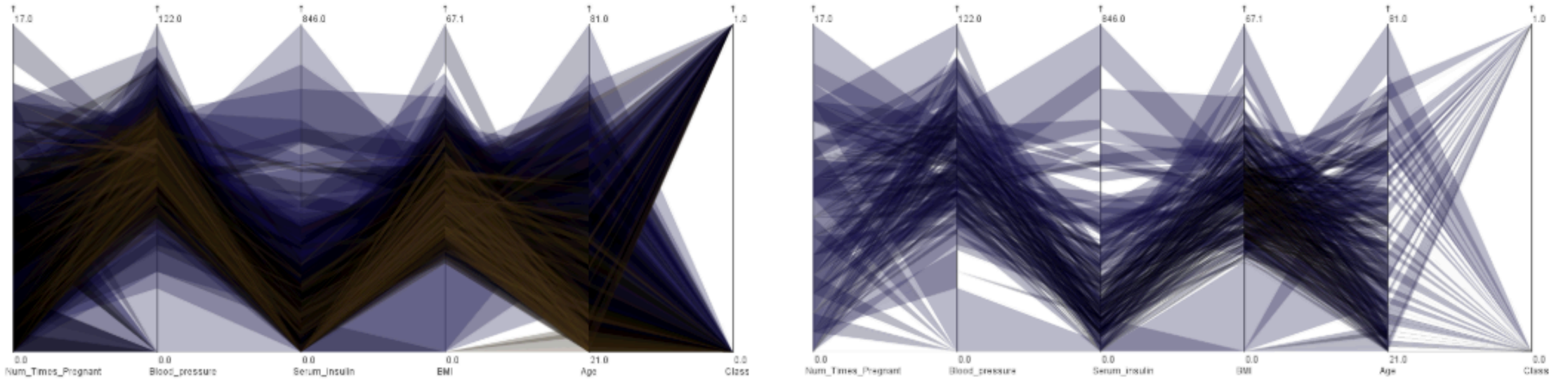


Figure 4: Multi-dimensional (left) and clustering by axis pair (right). Separate clustering for each axis pair creates smaller clusters that represent the structure of the data much better than clustering across all dimensions.

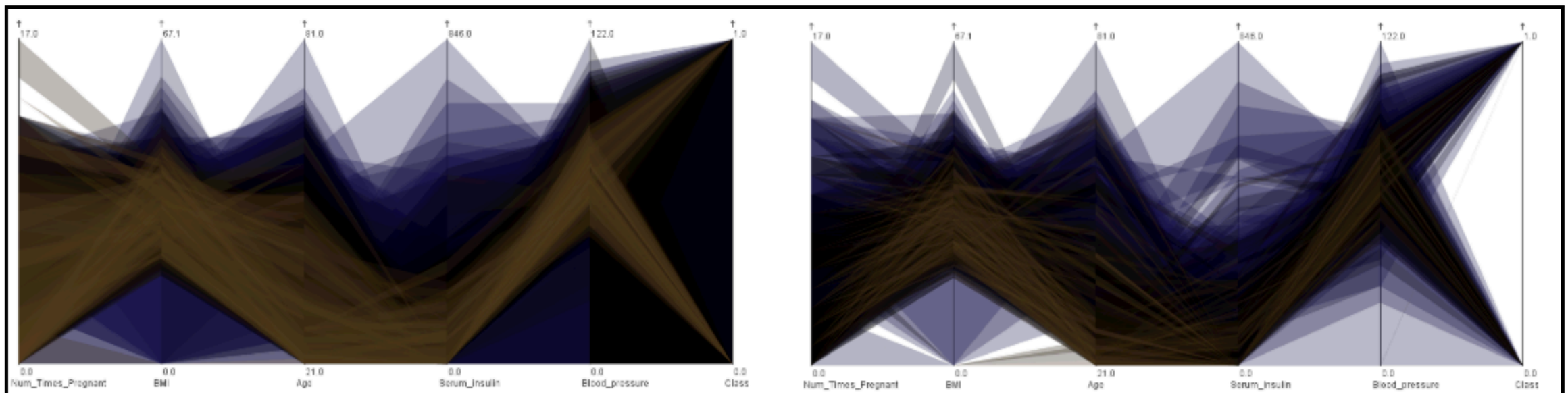


Figure 5: Cost functions used for clustering: Information loss (left) and Manhattan distance (right) used in multi-dimensional clustering. Manhattan distance leads to smaller clusters and more visible structures.

[Dasgupta2011b][DK11b]

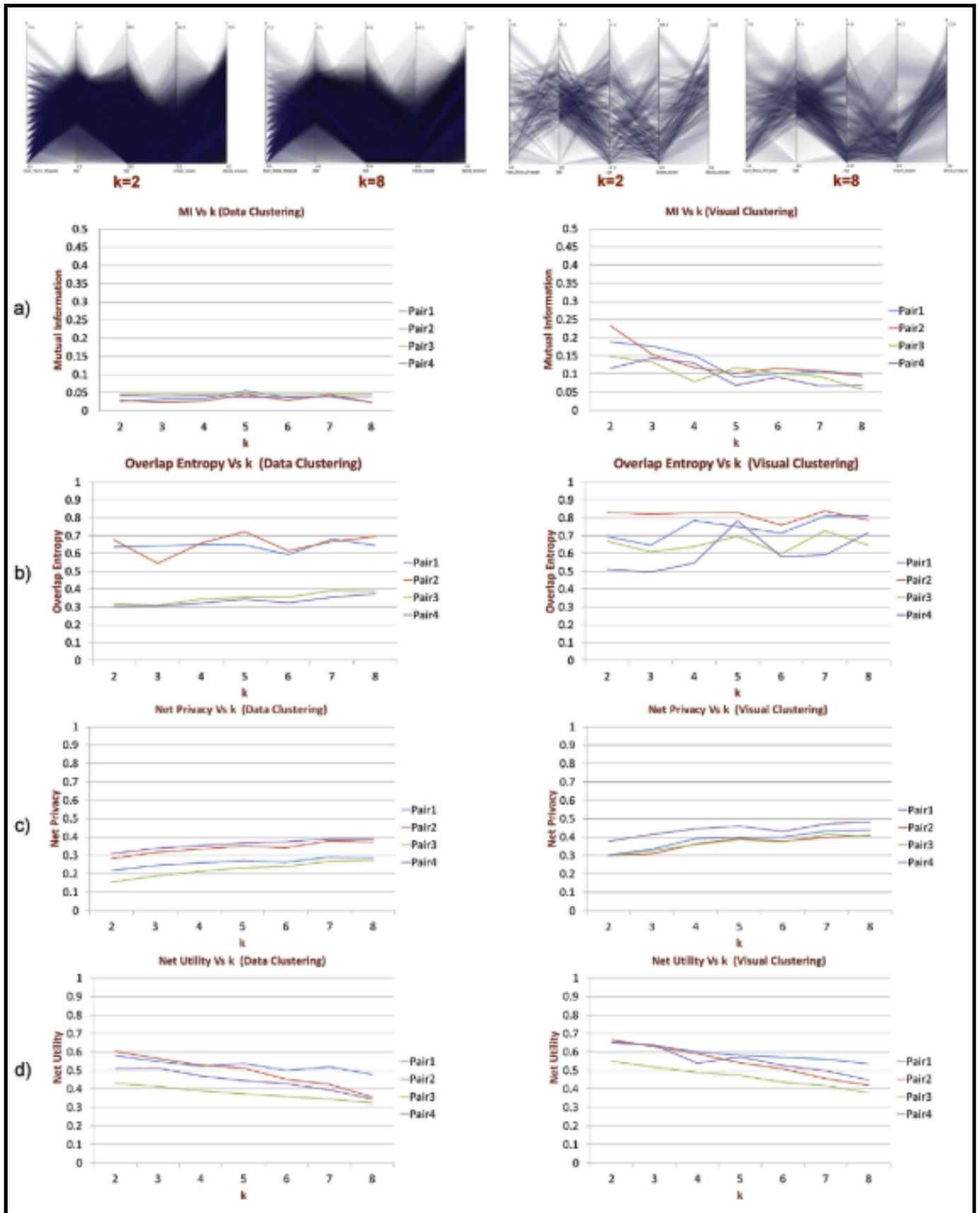
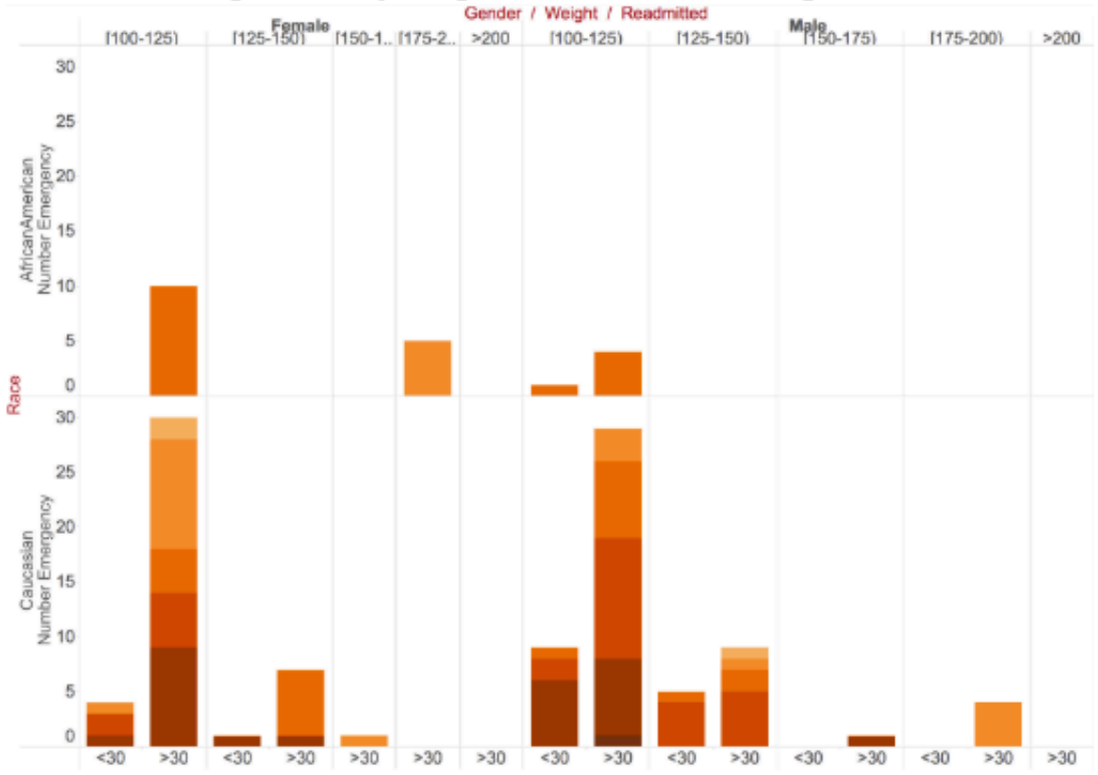


Figure 6: *Comparison of privacy and utility metrics for four different axis pairs in case of data clustering and visual clustering.*

Metrics for privacy and utility.

[Dasgupta2013][DCK13]

a) Bar chart representing hospital visits of diabetic patients



b) Treemap representation of the same data

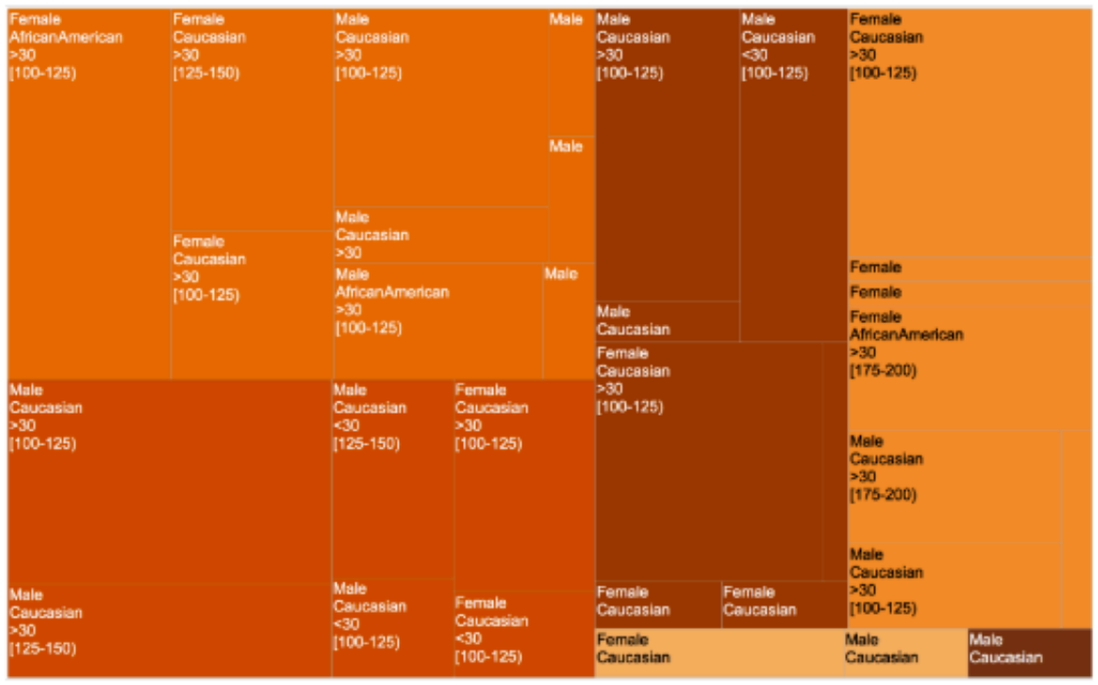


Fig. 2. Aggregate data visualizations such as bar charts (color mapped to age groups of patients) and tree maps (color mapped to age groups of patients and area mapped to their number of emergency visits) might be necessary but not sufficient to fulfill the conditions for privacy-preservation. In a) the bar chart reveals patterns like: for African American women having a high frequency of *readmitted* (> 30) and *highest number of emergency visits*, the only age group is 50 to 60 represented by dark orange. In b) the same information is revealed, but it does not pop out as much as in the bar chart. With the presence of externally available information, involving quasi-identifiers age and race, such patterns can be construed as both sensitive, and having a high potential to reveal individual identity.

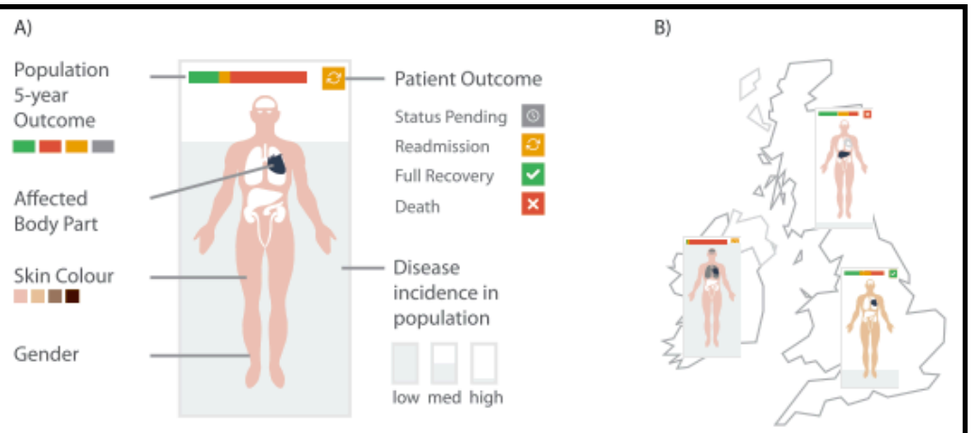


Fig. 4. A) An example glyph design showing some patient survival parameters for an unspecified disease. B) The glyphs are placed on to a map of the UK and Republic of Ireland to show three patients with differing diseases.

Examples showing data aggregation and binning is not enough.
Glyph example shows unsuspecting shapes can also lead to disclosure.

[Dasgupta2014][DMARC14]

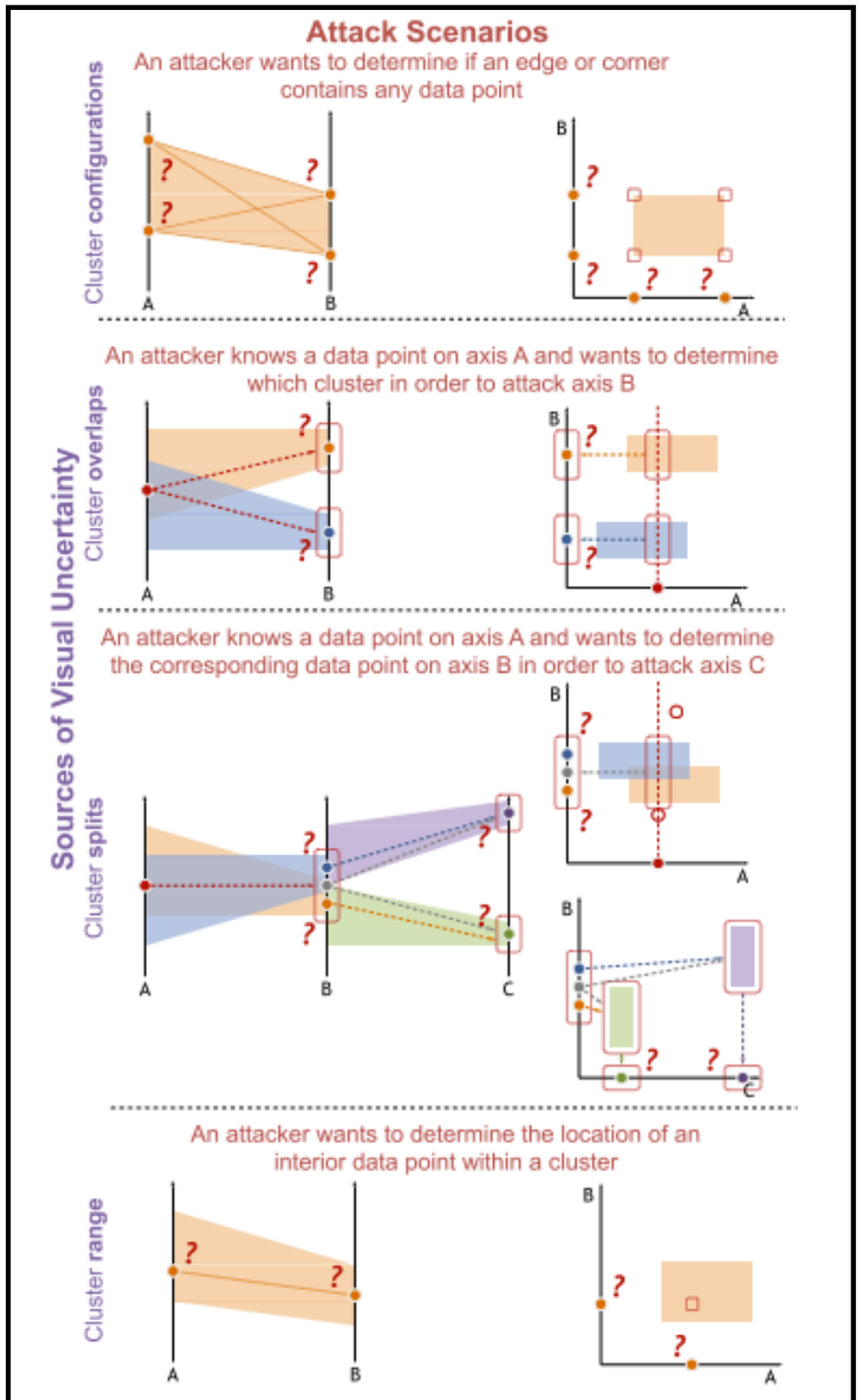


Figure 2: **Illustrating the relationship between adversarial attack scenarios and sources of visual uncertainty** in cluster-based parallel coordinates and scatter plots. Uncertainty due to cluster ranges and configurations helps in privacy-preservation when an adversary attempts to gain knowledge about the data at a lower level of granularity than what is shown, in case of **journalist re-identification scenario**. Uncertainty due to overlaps and split helps in privacy-preservation when an adversary attempts to determine the cluster membership of a known data point in case of the **prosecutor re-identification scenario**. These scenarios may not follow a particular sequence and one scenario can lead to another.

Attack scenarios and visual uncertainty.
A number of attack scenarios can be imagined based on which we calculate metrics for quantifying risks.

[Dasgupta2019][DCK19]

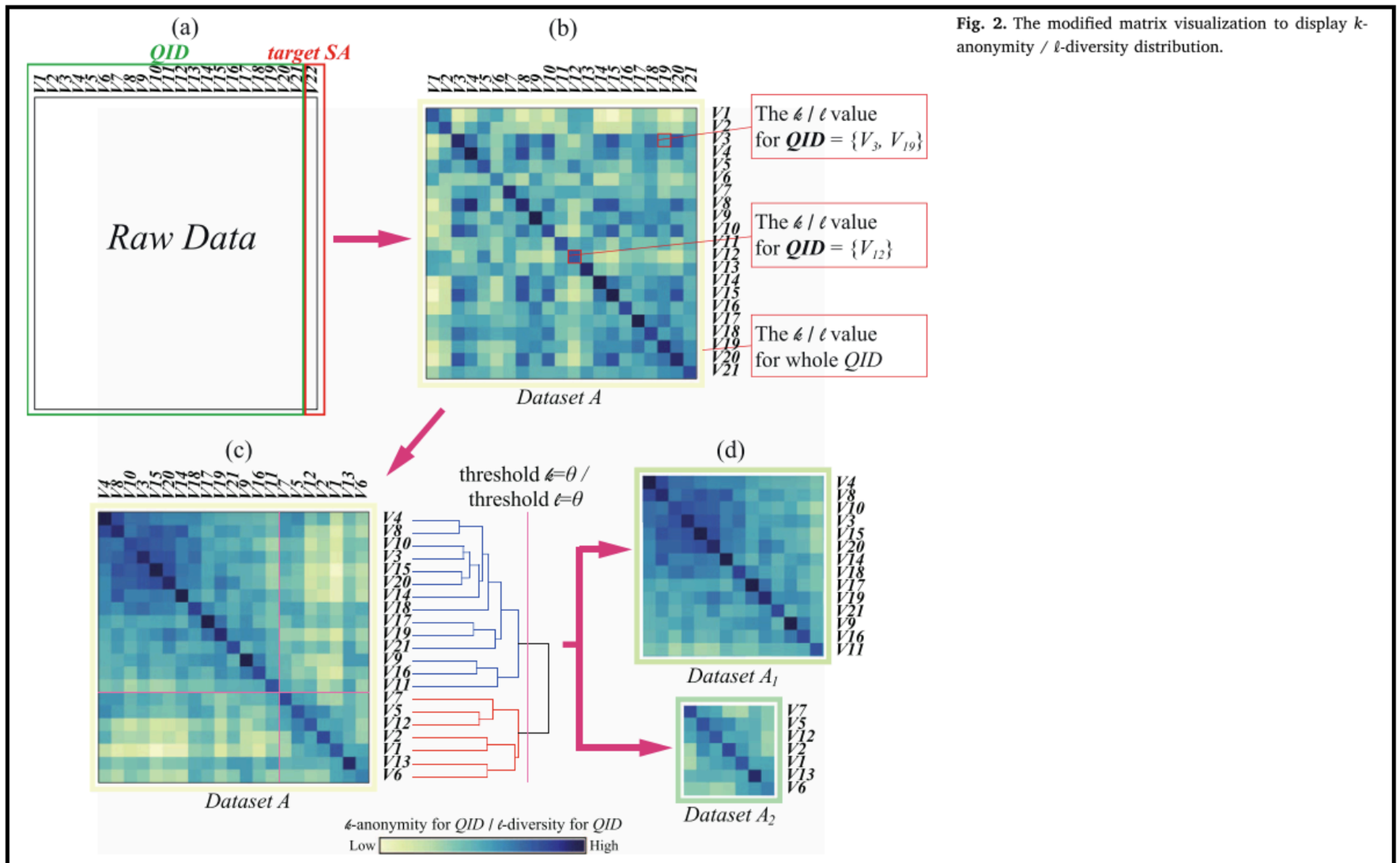


Fig. 2. The modified matrix visualization to display *k*-anonymity / *ℓ*-diversity distribution.

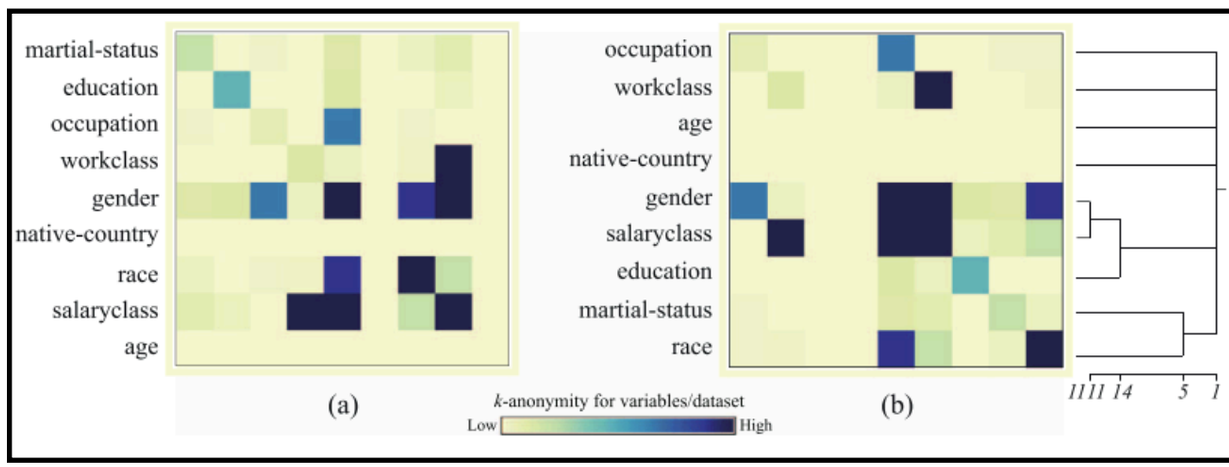


Fig. 3. A demonstration for how the ODD Visualizer works with regard to a benchmark.

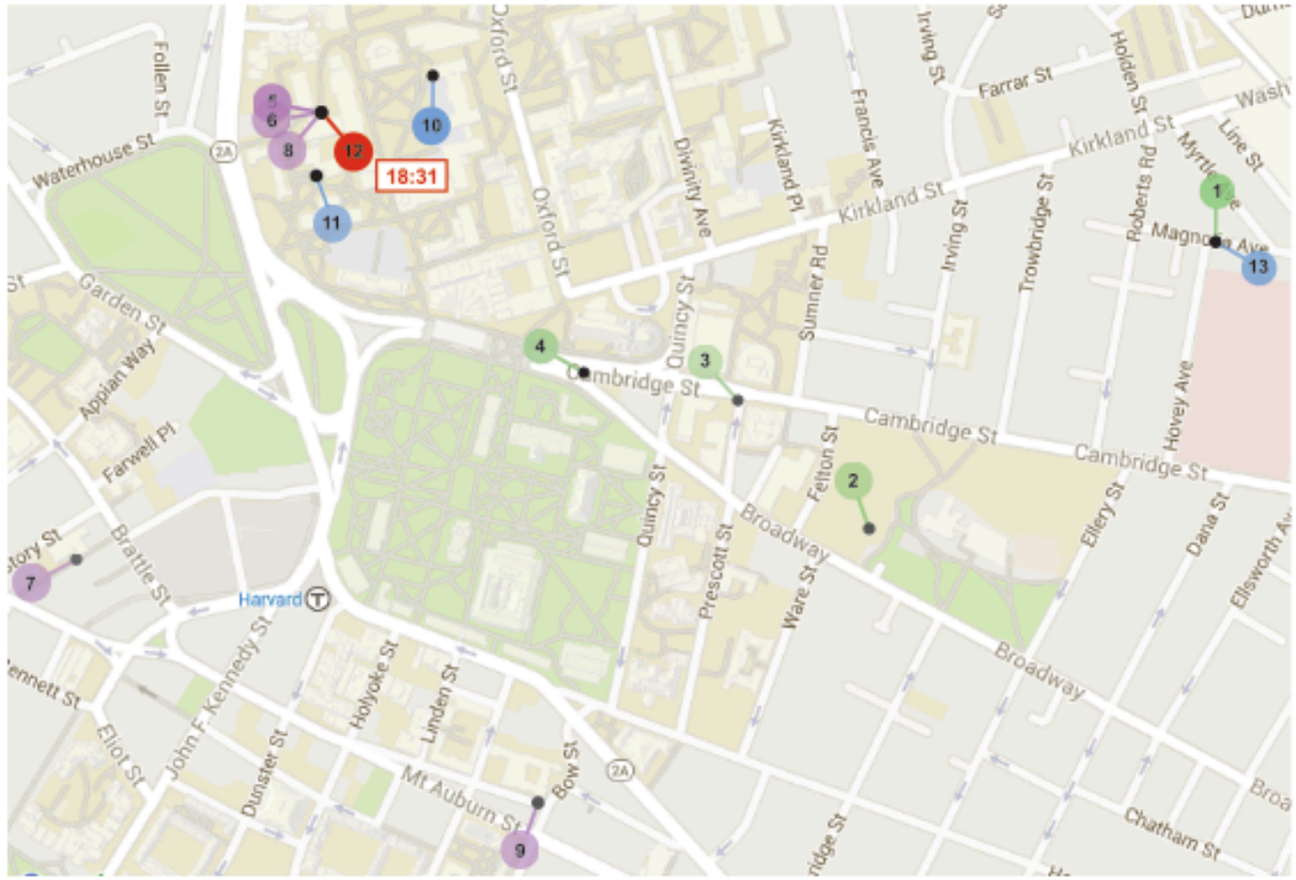
[Kao2017][KHC17] Privacy re-identification process are too cumbersome for the data owners and this gives rise to the privacy leakage in open datasets. This paper presents a novel visualization interface named ODD visualizer which will help in open data de-identification, i.e, if there is any privacy leakage in the dataset. It uses heat maps to display k-anonymity and l-diversity distribution.

(a) Textual (table-based)

ID	Latutude	Longitude	Address	Time
1	42.3769	-71.1073	Magnolia Ave.,Cambridge	7:25
2	42.3743	-71.1111	Broadway, Cambridge	8:52
3	42.3755	-71.1155	Cambridge St., Cambridge	9:19
4	42.3753	-71.1136	Cambridge St., Cambridge	9:25
5	42.3783	-71.1192	Massachusetts Av., Cambridge	12:01
6	42.3783	-71.1193	Massachusetts Av., Cambridge	12:30
7	42.3737	-71.1220	Brattle St., Cambridge	13:30
8	42.3783	-71.1193	Massachusetts Av., Cambridge	14:12
9	42.3717	-71.1154	Bow St., Cambridge	15:00
10	42.3784	-71.1171	Oxford St., Cambridge	17:35
11	42.3776	-71.1190	Massachusetts Av., Cambridge	17:49
12	42.3782	-71.1192	Massachusetts Av., Cambridge	18:31
13	42.3769	-71.1073	Magnolia Av., Cambridge	19:33

Location in question Morning Afternoon
Evening End of morning End of afternoon

(b) Visual (map-based)



ID	Time
1	7:25
2	8:52
3	9:19
4	9:25
5	12:01
6	12:30
7	13:30
8	14:12
9	15:00
10	17:35
11	17:49
12	18:31
13	19:33

Location in question Morning Afternoon
Evening End of morning End of afternoon

Figure 3. Textual (a) and visual (b) representations used in the study. The location density displayed is *Low* (1 day). The detailed captured time is shown in addition to the color depending on the period of the day.

[Liccardi2016] [LARC16] This paper uses a novel geographical map visualization in order to infer places of interest (like home, work, social places etc.) of a person from geo-location enabled tweets. This paper also conducts an empirical study to understand how accurately these places of interest can be analyzed from a text based interface and a visual map based interface. It also discusses the implications of the findings of this study through a privacy perspective.

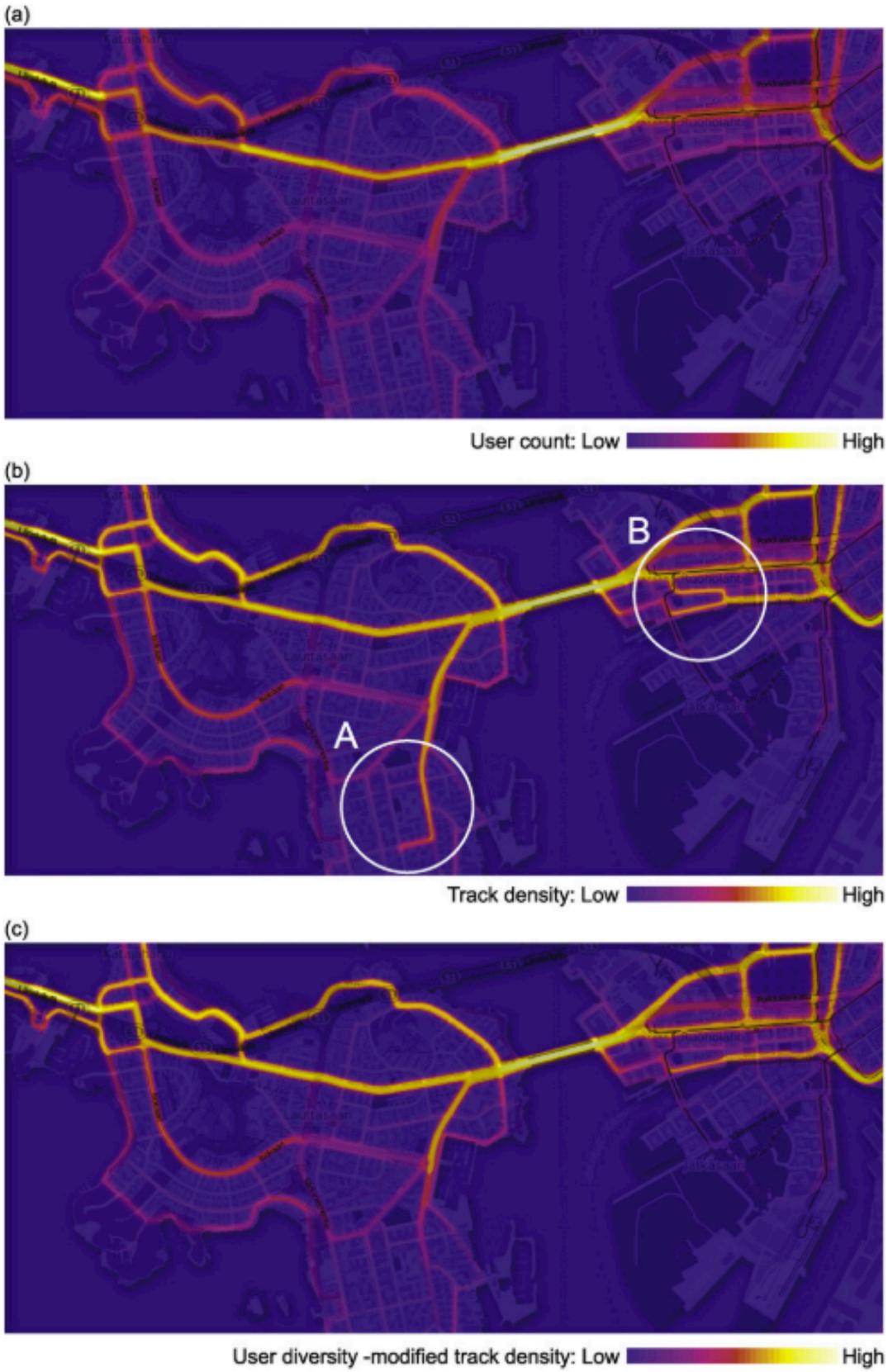


Fig. 4. Heat maps based on (a) privacy-preserving user count calculation (ppUCC), (b) privacy-preserving kernel density estimation (ppKDE), and (c) privacy-preserving kernel density estimation modified with the user diversity index (ppDIV). Major differences between the methods are found in the highlighted regions A and B. The map contains data from the Topographic database by the National Land Survey of Finland, 7/2012; © OpenStreetMap contributors.

[Oksanen2015][OBSW15] This paper has used movement data from mobile sports tracking applications to generate privacy-preserving heat maps in which the trajectories and the diversities of the users can be studied. This will also help to reduce the bias caused by participation inequality in this type of study. This paper generated heat maps from public cycling workouts data and compared it ot privacy-preserving kernel density estimation (ppKDE) and privacy-preserving user count calculation (ppUCC)

Pair	ID	First name	Last name	DoB(M/D/Y)	Sex	Race	Choice Panel
1	1990443570	BOYLE	JASON	11/14/1980	M	W	H M L L M H Different Same
	1990443570	JASON	BOYLE	11/14/1980	M	W	
2	1000027594	CHARLES	GREEN	07/10/1930	M	W	H M L L M H Different Same
		CHARLES	GREEN	07/10/1903	M	W	

Figure 1. Above are two pairs of data records in the study application in the *baseline* condition with all information visible.

Pair	ID	FFreq	First name	Last name	LFreq	DoB(M/D/Y)	Sex	Race	Choice Panel
1	✓	①	&&&&&	@@@@@	•••	✓	M	✓	H M L L M H Different Same
	✓	∞	@@@@@	&&&&&	•••	✓	M	✓	
2	**********	∞	✓	✓	2.5	07/10/1930	M	✓	H M L L M H Different Same
	?	∞	✓	✓	2.5	07/10/1903	M	✓	

Figure 2. Example from the study application showing supplemental markup and value masking. The two pairs in this example are shown in the *moderate* condition using the same pairs shown in Figure 1. The visual markup highlights discrepancies, provides information about name frequency, and hides common values.

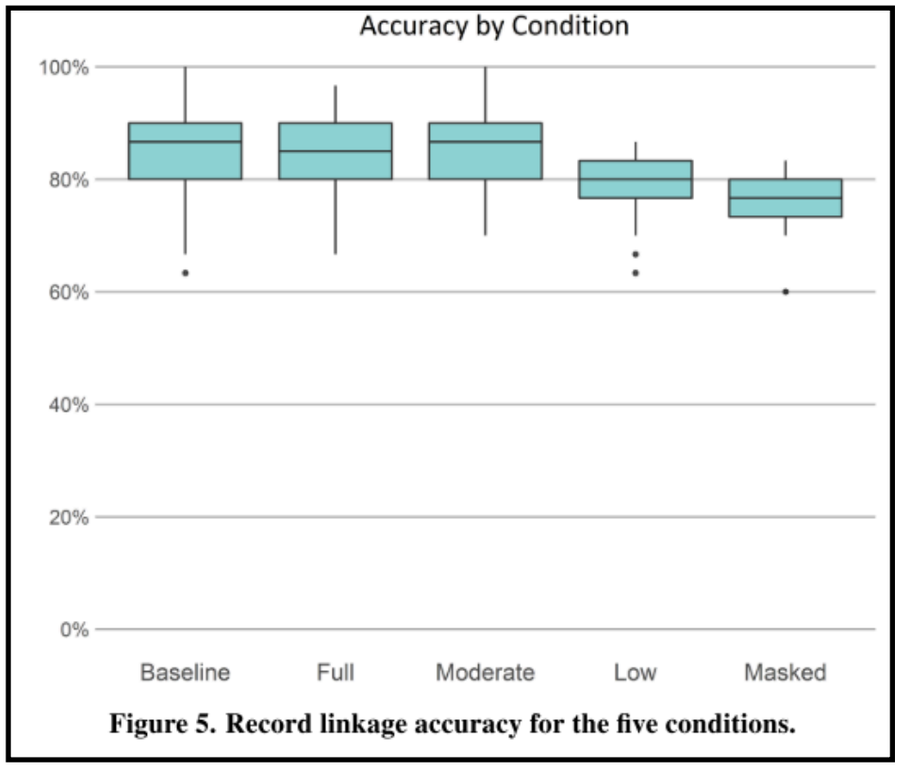


Figure 5. Record linkage accuracy for the five conditions.

[Ragan2018][RKIW18] This paper presents an interactive interface where the user starts with fully masked de-identified data and later clicks to open when more information is required for making better decisions. This is a system that reduces privacy risk through on-demand incremental information disclosure. Box plots have been used to analyze the test results in different masking levels like full, moderate, low and masked.

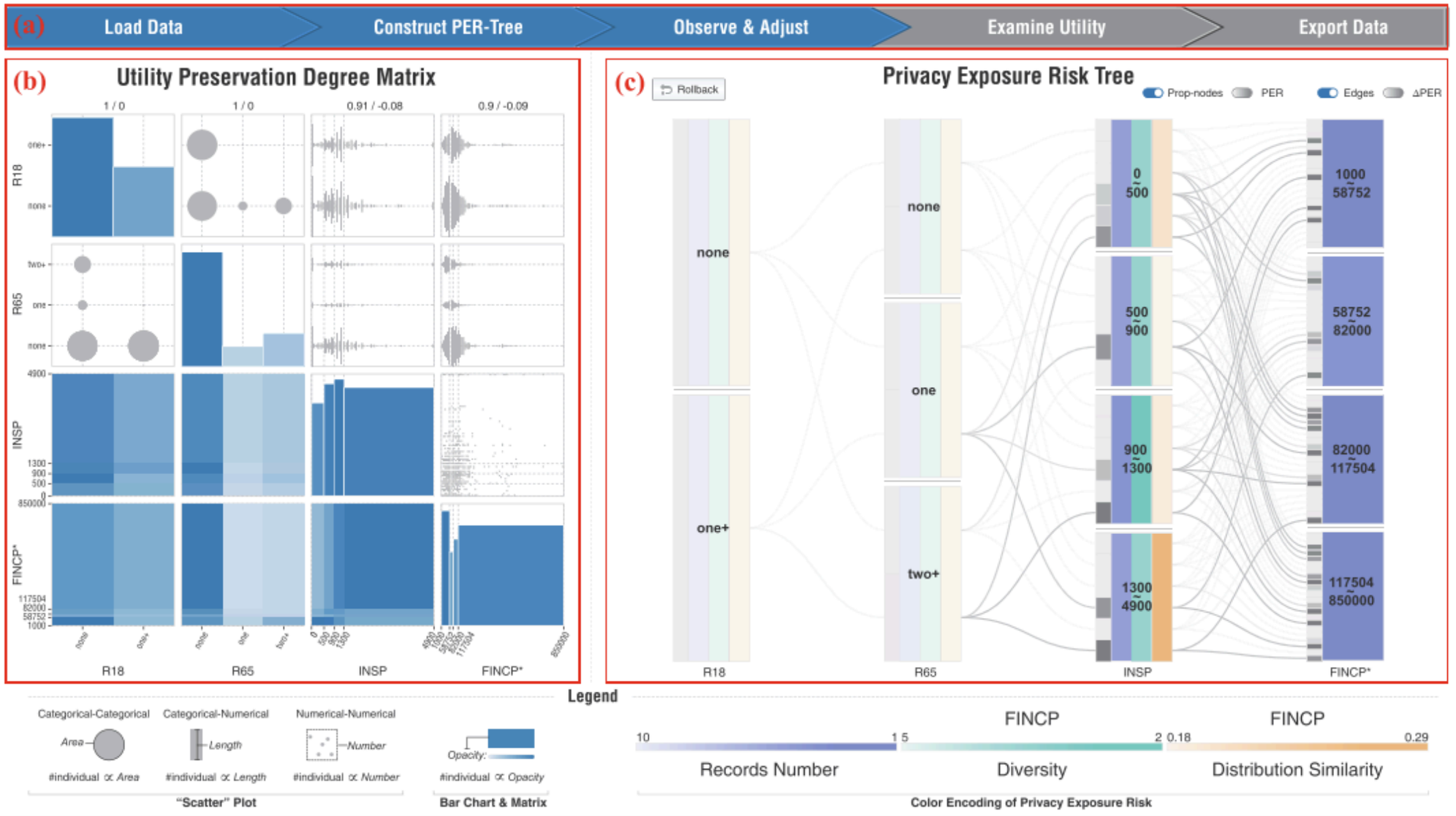


Fig. 1. Our utility-aware visual data-anonymizing process follows (a) a 5-step pipeline and is facilitated with two main visualization components: (b) utility preservation degree matrix (UPD-Matrix) and (c) privacy exposure risk tree (PER-Tree). The PER-Tree helps our users identify privacy issues in the underlying data and provides interactions to address the detected privacy issues. The UPD-Matrix presents the difference between the processed data and the original data. Users can use the chart to examine how utility of data changes during the anonymization process.

One of the few examples of a tool for evaluation of trade-offs.

[Wang2017][WCC17]

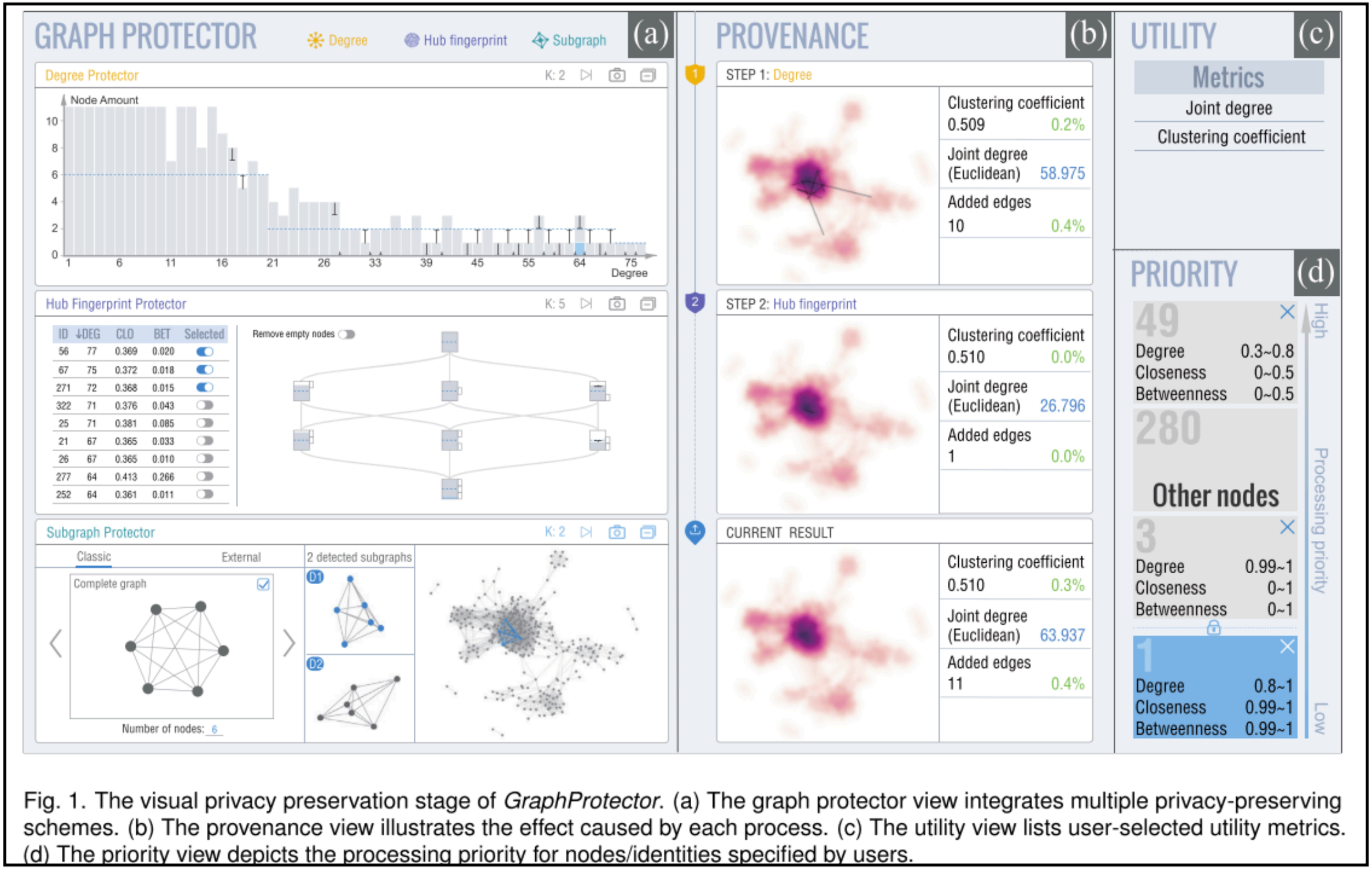


Fig. 1. The visual privacy preservation stage of *GraphProtector*. (a) The graph protector view integrates multiple privacy-preserving schemes. (b) The provenance view illustrates the effect caused by each process. (c) The utility view lists user-selected utility metrics. (d) The priority view depicts the processing priority for nodes/identities specified by users.

GraphProtector is a great example of a tool for algorithm comparison.

[Wang2018a]WCC18

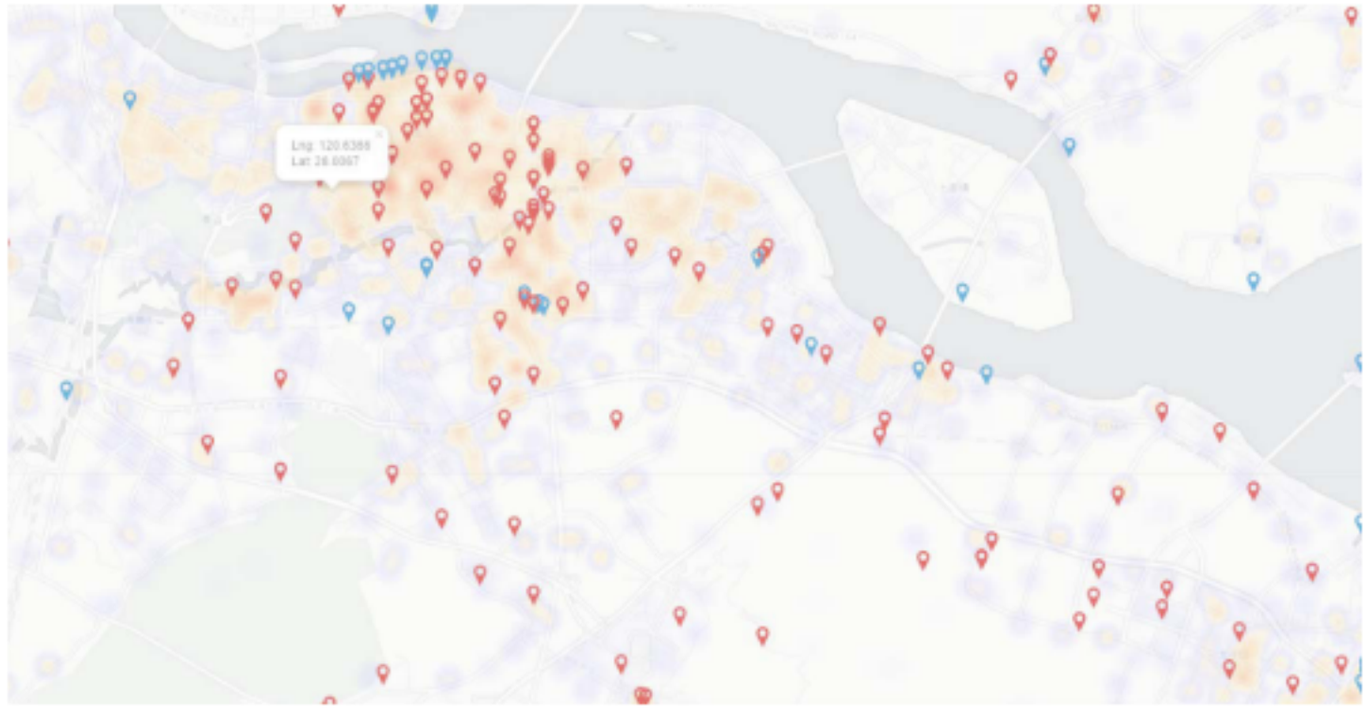


Fig. 5. The main view for the locating task. The entire map with all the information that can be found in the road network, heatmap and POI features are shown.

[Wang2018b]WGL18 This paper uses a novel geographical map visualization in order to infer places of interest (like home, work, social places etc.) of a person from geo-location enabled tweets. This paper also conducts an empirical study to understand how accurately these places of interest can be analyzed from a text based interface and a visual map based interface. It also discusses the implications of the findings of this study through a privacy perspective.

No Visualization

[Gkoulalas-Divanis2014][GLS14] This paper talks about preserving privacy while publishing electronic health records. Also talks in details about how to prevent identity disclosure, membership disclosure and attribute disclosure. This paper also discusses how privacy-preserving data sharing can also be facilitated in the non-interactive privacy scenario. This scenario assumes that the data are deposited into a secure repository and can be queried by external data users. But the authors also note that complex queries are difficult to support in the interactive scenario (similar to statistical databases) and several analytic tasks like visualization, require individual records, as opposed to aggregate results or models. Hence, they have noted that data publishers need to carefully select the appropriate privacy-preserving data sharing scenario based on their needs.

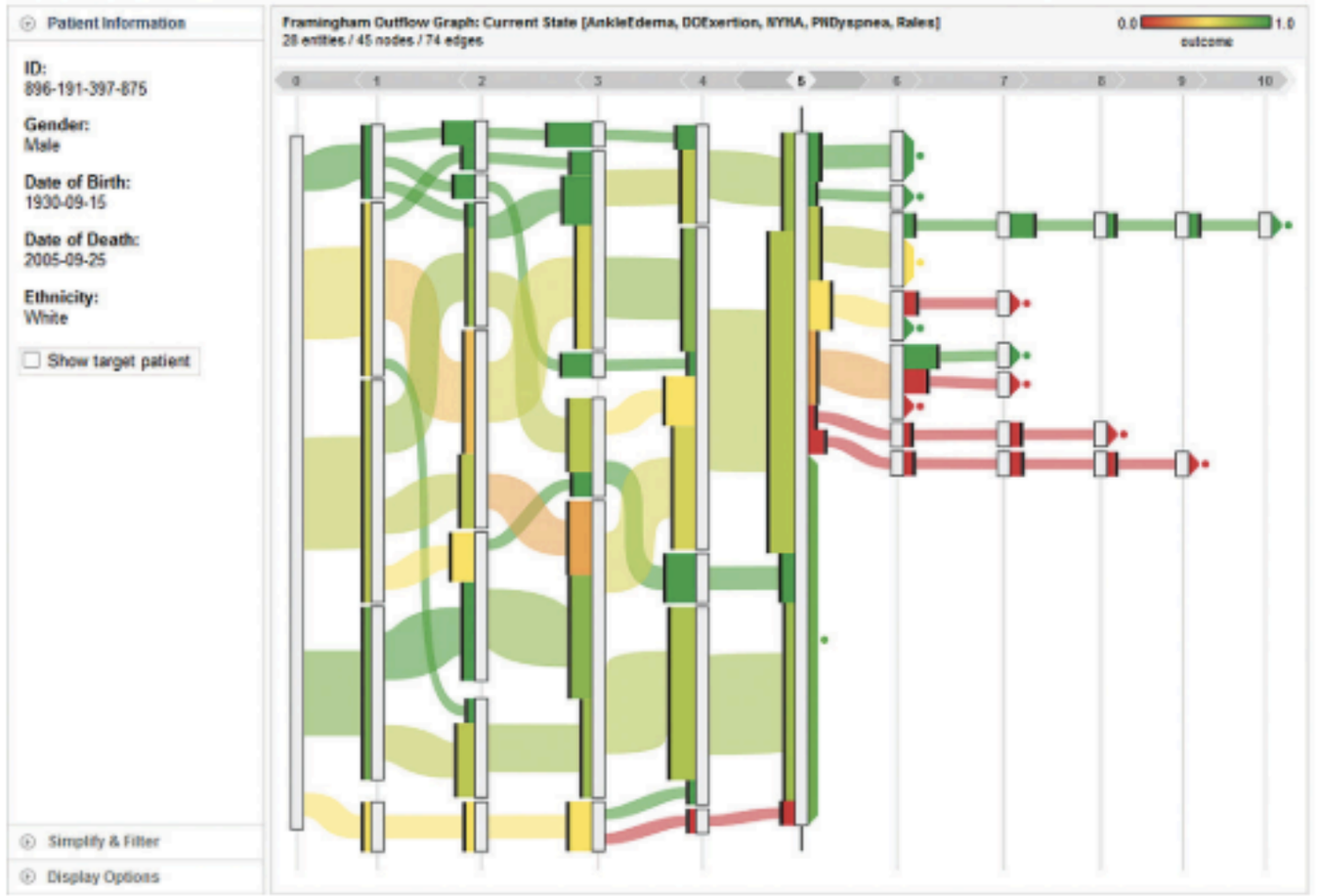


Figure 4. Flow-based visualization methods have been used in several different projects to analyze temporal event data, including the Outflow visualization shown here.[12] This example shows variations in the order of symptom onset for a cohort of heart failure patients and the medical outcomes associated with each subgroup.

[Gotz2016][GB16] This paper presents a novel visualization method to analyze the user's temporal medical data. Flow based visualization has been used to represent the temporal data like the order of symptoms onset for heart failure patients. The main challenge lies in limiting the amount of information for privacy issues and this may significantly hamper the process of record linkage in this medical data.

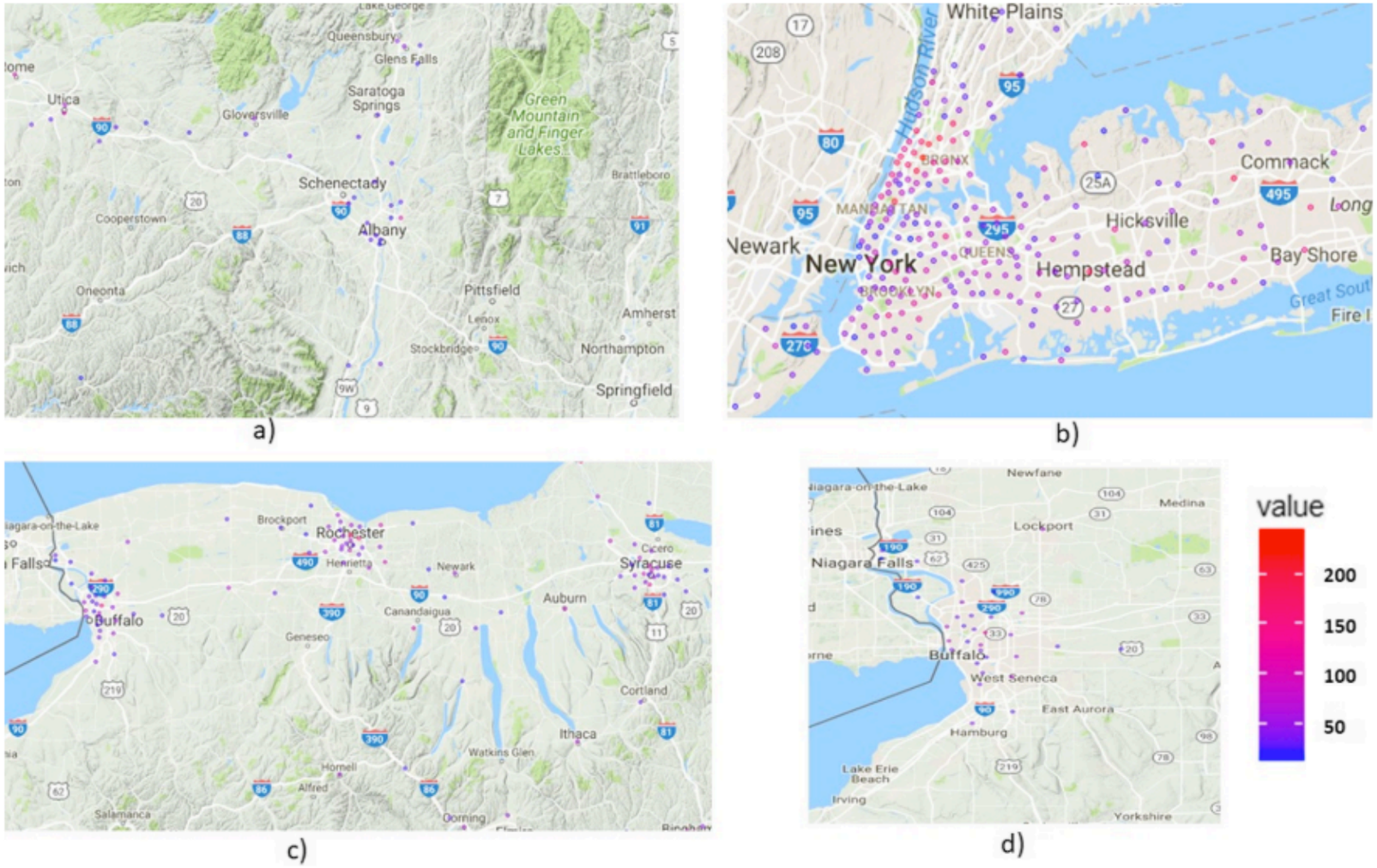


Fig. 3. Heatmaps of (a) Albany area, (b) NY City area, (c) North side of NY state, (d) Buffalo area. Heatmaps show that the distribution of hospitalized flu patients by the zip code in period between 2003 and 2012, was highly concentrated in five big cities (Albany, Buffalo, New York City, Rochester and Syracuse). Heatmaps show that the routes of distribution follow highways (highways 81, 86 and 90, in Albany area highways 87 and 9 and in Buffalo area, highway 190 toward Niagara Falls).

[Ljubic2019][LGGPO19] This paper uses geographical heatmaps to present the distribution of influenza in a certain area. This helps in finding the affected area in a certain geographical region which may be helpful to healthcare officials. A privacy leakage in these geographical heatmaps may allow the identification of certain patients, leading to identity disclosure.

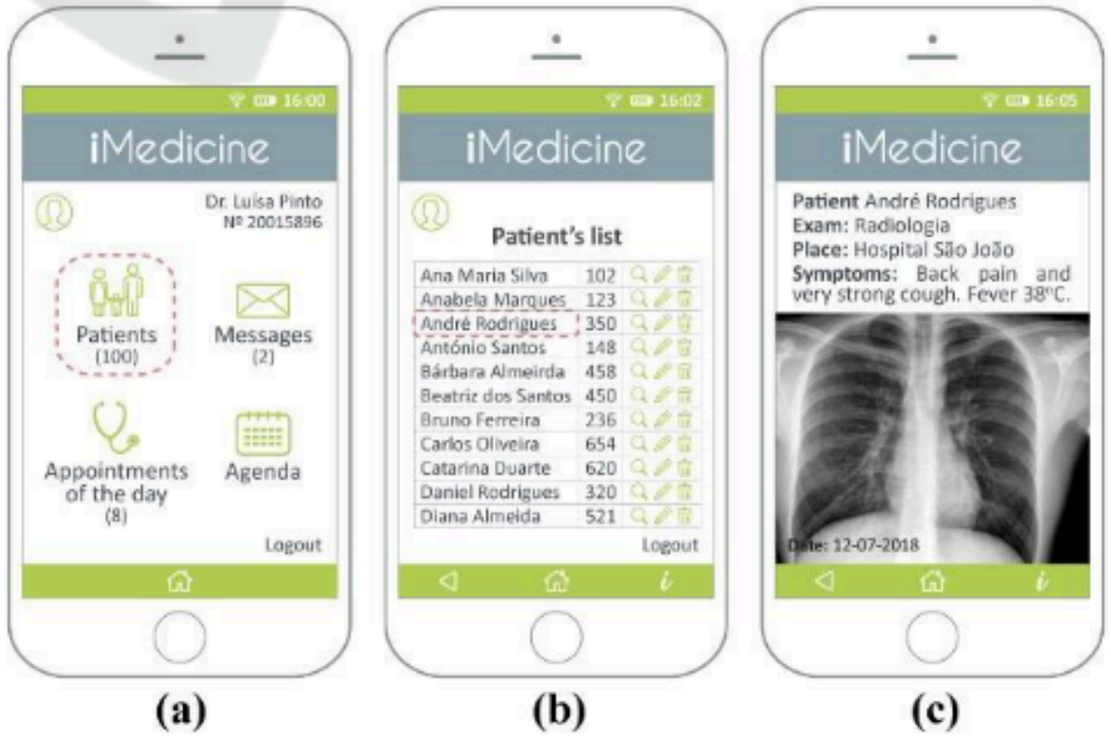


Figure 3: Before using the AGVI Dr. Luísa is able to see everything about her profile, patient's information, messages, appointments and agenda.

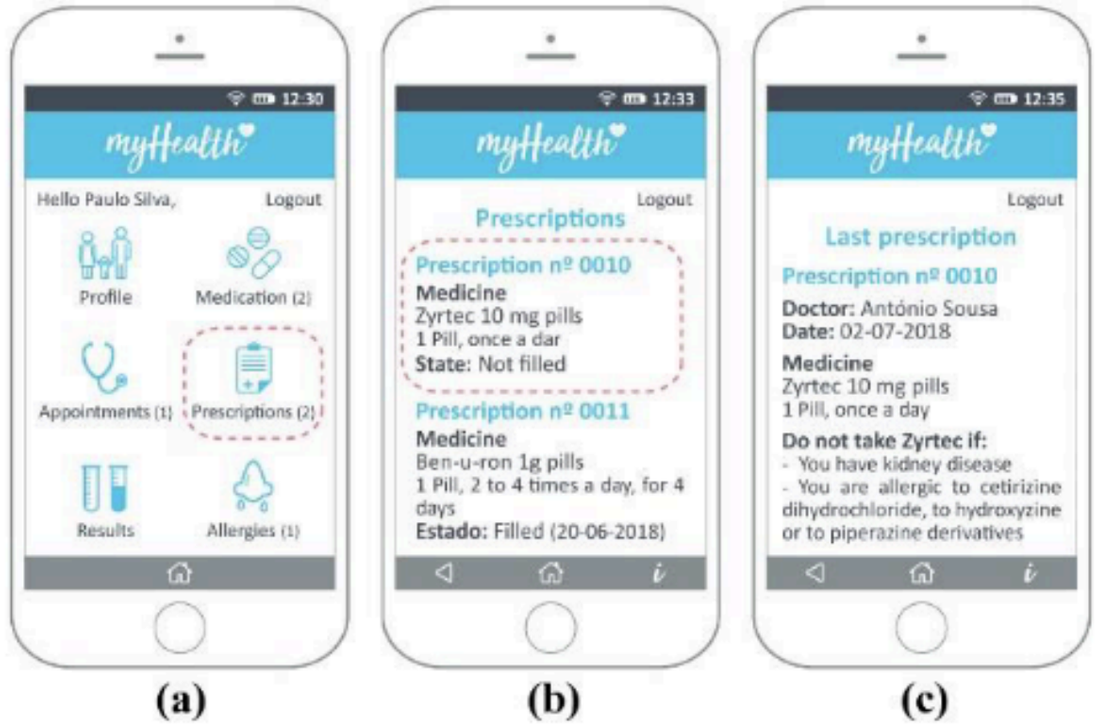


Figure 2: After using the AGVI the app shows information according with user's characteristics and needs with improved visual security.

[Muchagata2019][MVF19] This paper presents a text-based interface in a mobile application which will help patients and healthcare professionals to monitor health data. The most important feature of this visualization, named Adaptive Graphical Visualization Interface (AGVI), is the interface is user-adaptive, i.e., it changes according to the user's needs. This paper observes that adaptive visualization techniques can influence the users' perspective on security and privacy of a mobile application but the roles of the user (patient or healthcare professional) and their goals (searching for medications or analyzing patients' tests) can influence this perspective.

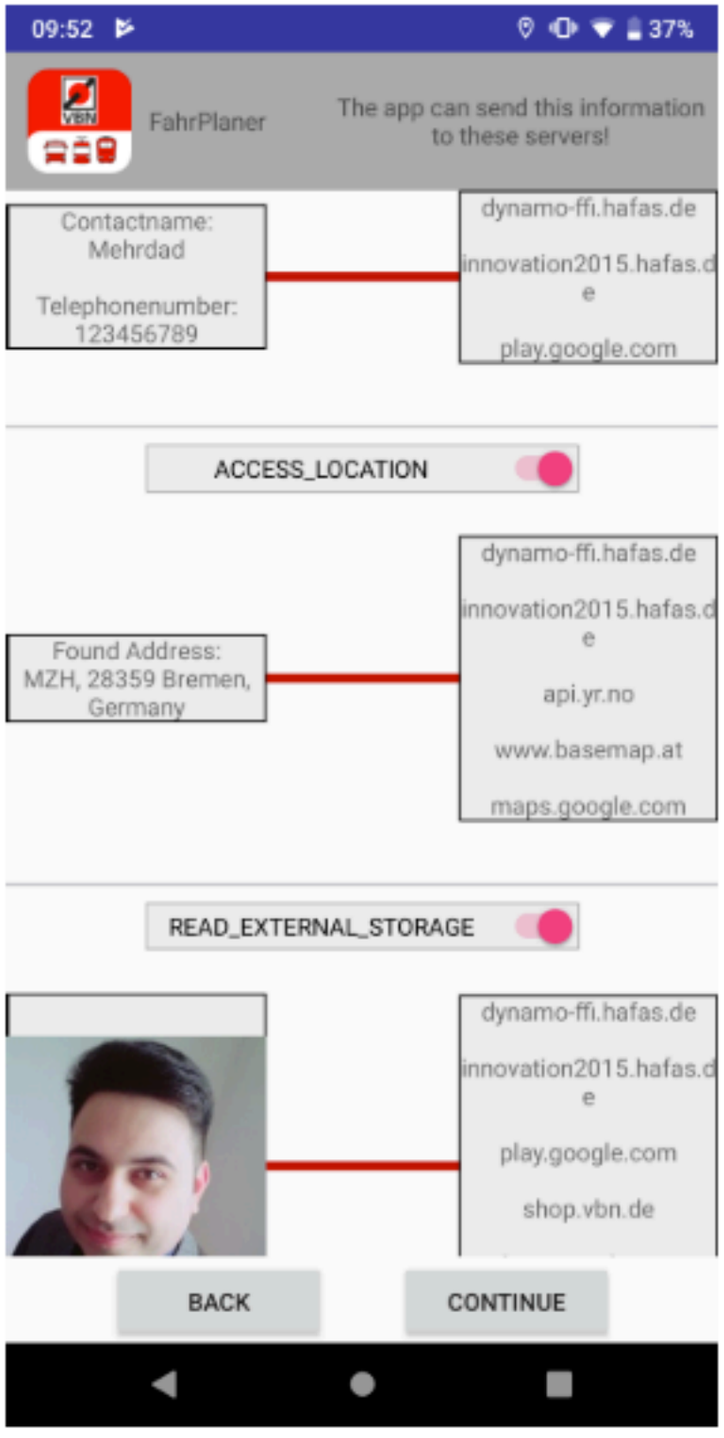


Figure 4: HappyPermi+Flow: For each permission, the left box indicates private user data that is accessible through granted permissions and the right box shows its destination in the form of URLs. This feature is available in the HappyPermi+Flow version.

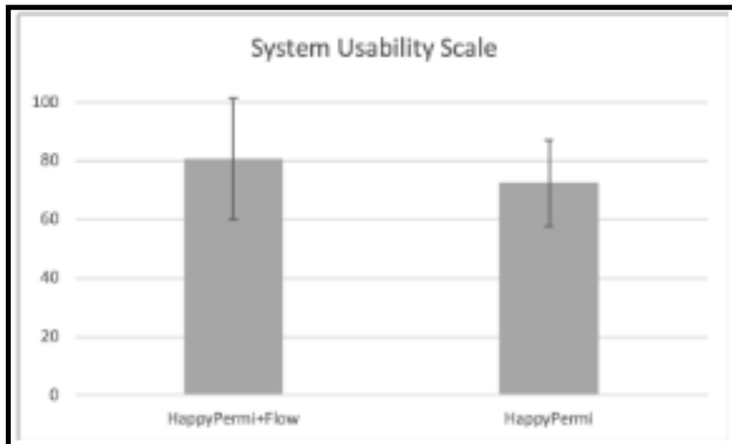


Figure 5: The SUS for both groups: The HappyPermi and the HappyPermi+Flow application

[Bahrini2019][BMMWS19] This paper discusses how a mobile application can help users to understand which user information is accessible by the granted permissions. This interactive visualization will help the users make an informed decision about whether to install a certain application or not. The authors claim that the results of their evaluation state that by promoting user awareness regarding permissions required by mobile applications (Android), users pay more attention to these permissions. The paper also tested system usability using error bars for different versions of the application and concluded that the version with more detailed description/flow of permissions has greater usability.

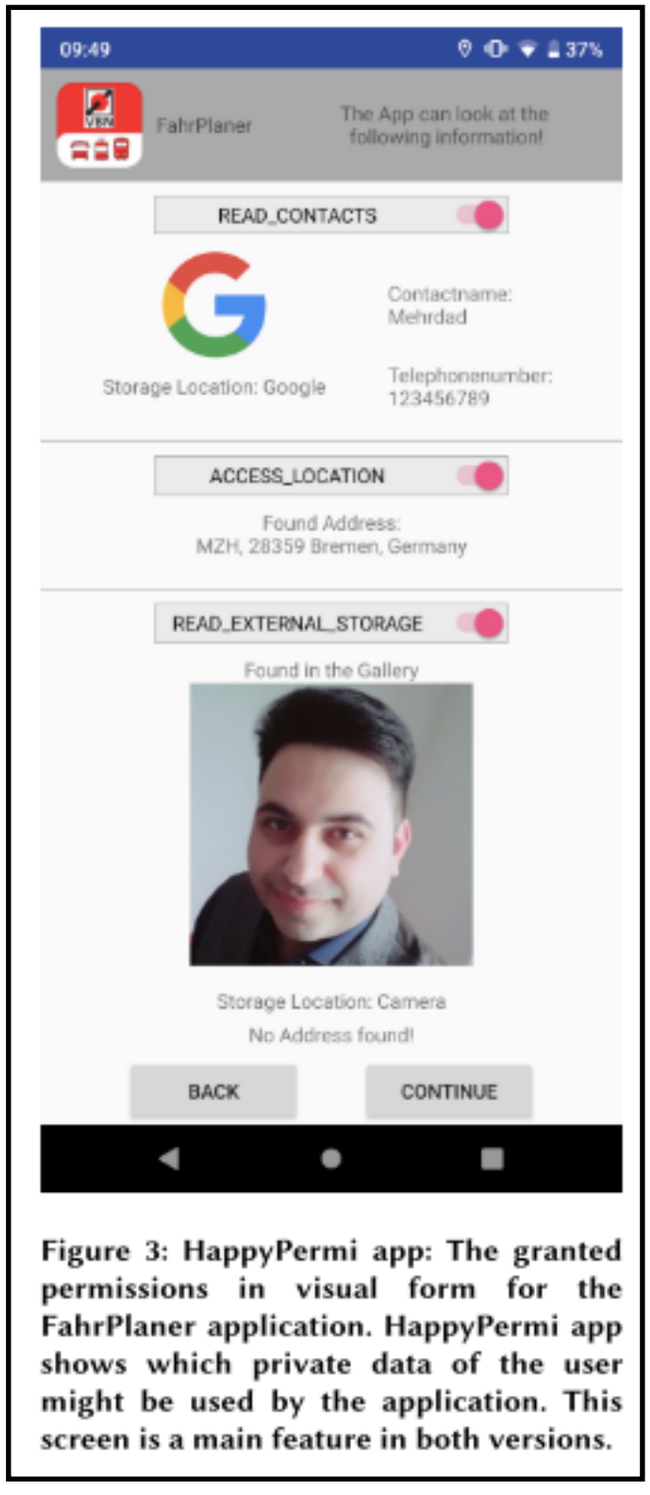


Figure 3: HappyPermi app: The granted permissions in visual form for the FahrPlaner application. HappyPermi app shows which private data of the user might be used by the application. This screen is a main feature in both versions.

[Bahrini2019][BMMWS19] This is the HappyPermi app without the detailed flows.

No Visualization

[Conti2005][CAS05] Since information visualization systems are an effective way to understand large amounts of data, they are prone to manipulation attacks. Hence the authors present a framework for information visualization system security analysis, a taxonomy of visualization attacks and technology independent principles for countering malicious visualizations. The authors talk about different attack techniques like “Cry Wolf” attack, displacement attack among others. This paper suggests that in order to protect from these attacks, the user needs to be educated, data generation and data flow need to be protected and the systems should be designed to protect the user, along with the consideration that the adversary is intelligent and well-informed.

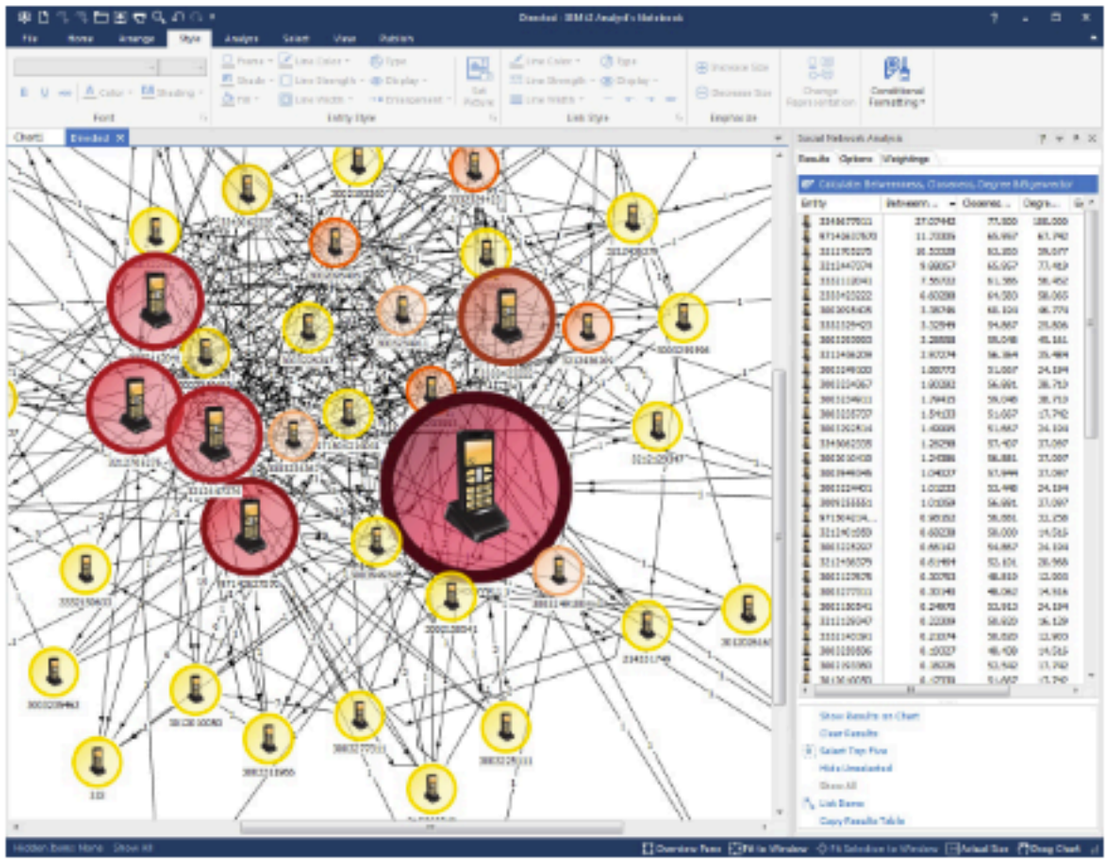
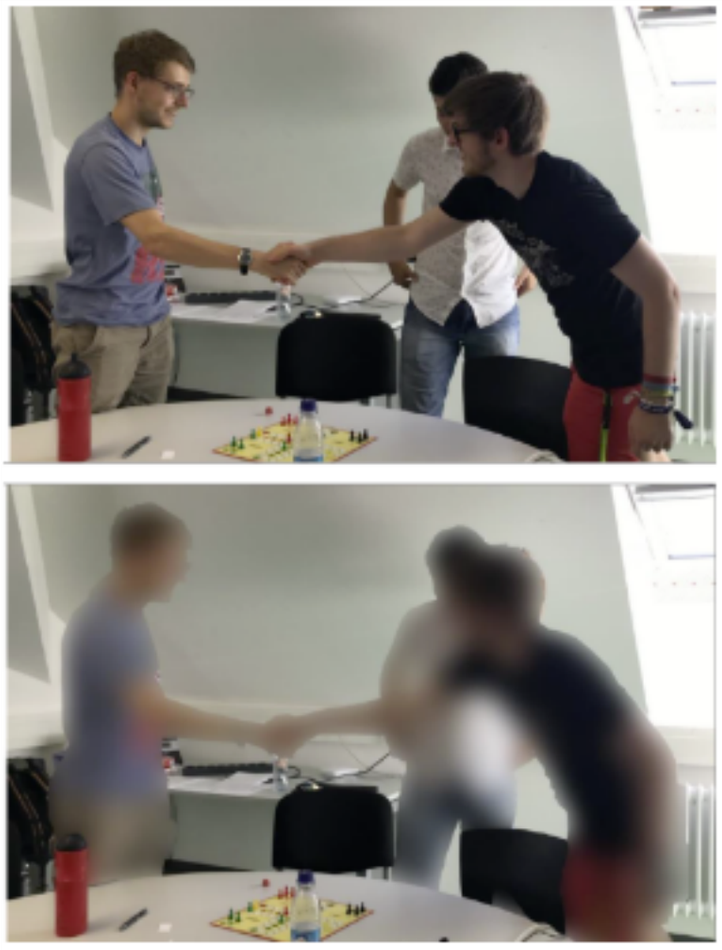
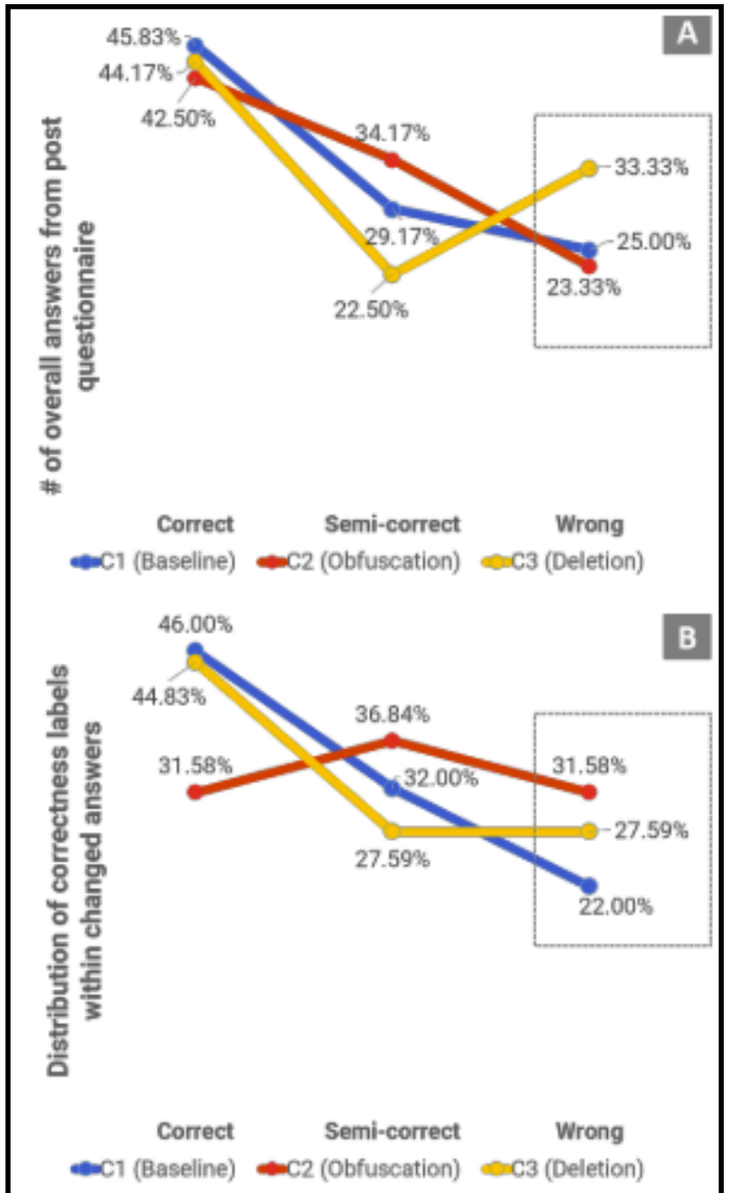


Figure 2: Example screen shot of a link chart made with IBM's i2 Analyst Notebook. Image courtesy of IBM

[Deeb2019][DEB19] This is a view of the IBM's i2 Analyst Notebook. It is used by analysts in law enforcement cases to visualize social networks. These visualizations were used to explain the insights developed in the case to the co-workers and prosecutors. The authors note that this is an interesting and challenging problem area for information visualization practitioners as the data in the human trafficking cases is collected from a wide range of sources, ranking from semi-structured databases to coded transcriptions of interviews. The authors also note that it is quite challenging to visualize the data with computational or analytic models with inherent uncertainty, but it is also difficult to create interactive interfaces that will allow investigators or law enforcement officials to make well-informed decisions.



Sidebar 1: Original photo of three participants (top) vs. obfuscated version (bottom) where bodies are blurred (gaussian blur, radius = 40 px). We found that obfuscated lifelogs allow viewers to remember more details but with less accuracy.



Sidebar 5: Overview of the correctness level for each condition (C1-C3). (A) Obfuscation of persons was associated with the highest number of *semi-correct* overall answers, while deletion of photos was associated with the highest number of *wrong* answers. (B) However, reviewing the answers using obfuscated photos (C2) resulted in the lowest ratio of *correct* answers and the highest ratio of *semi-correct* and *wrong* answers.

[Elagroudy2019][EKMIBS19] This paper talks about altering photos captured through lifelogging cameras, using visualization techniques like obfuscation. The authors conducted a study of two sessions where lifelog was generated in one session and they were reviewed in the other session. The first image shows an image and its obfuscated version. The study concluded that obfuscation improved the number of proper recalls by the participants in later session but did not improve the accuracy, as depicted by the line graphs. As mentioned in the paper, privacy-aware obfuscation in photos is commonly used in applications like Google Street View.

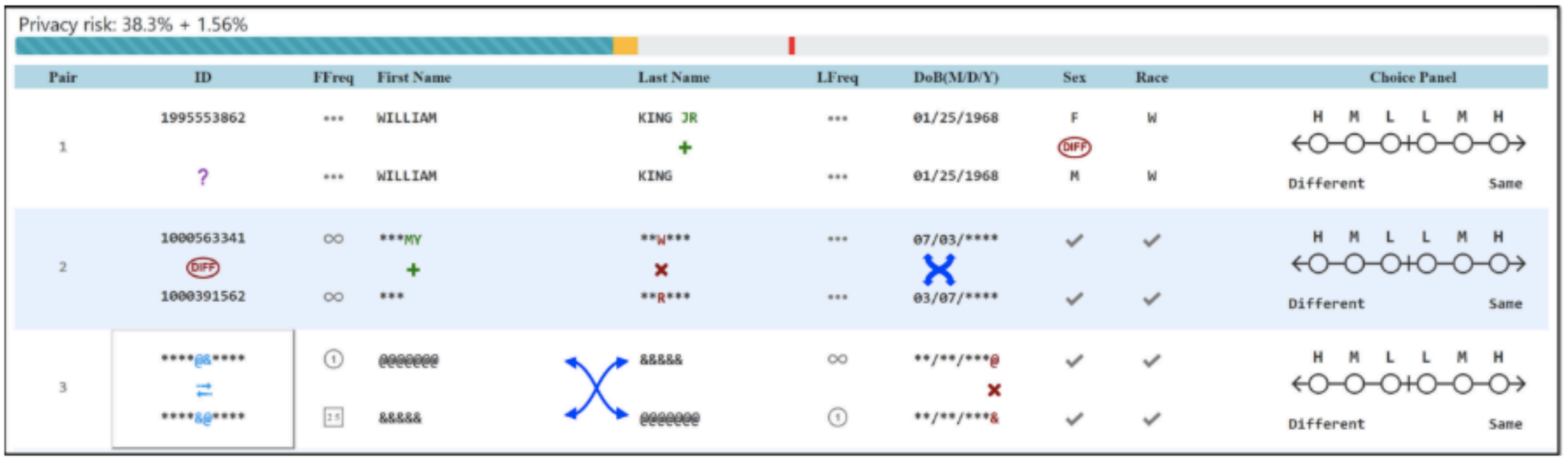
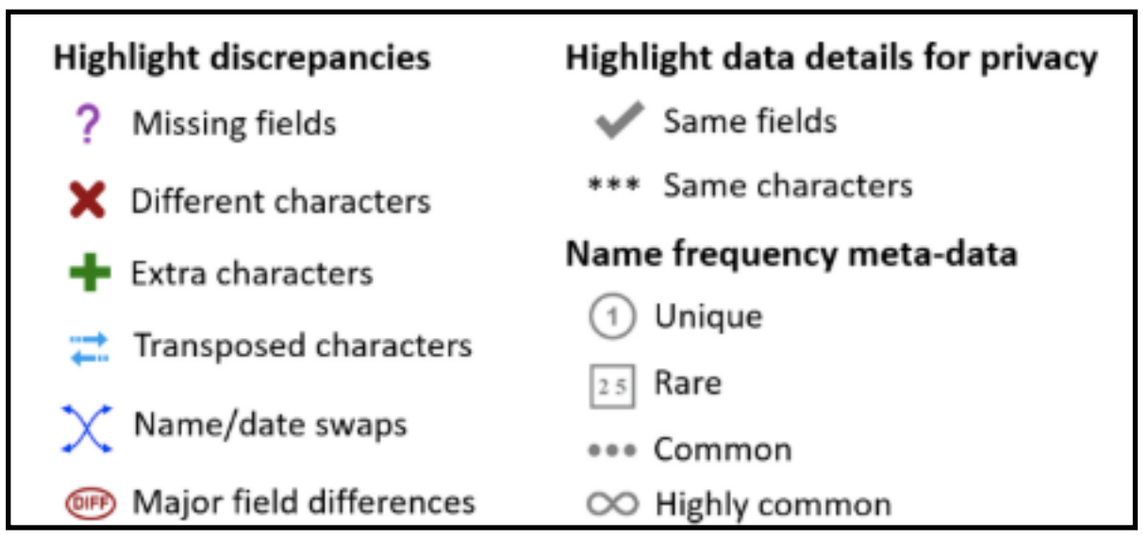


Figure 2: Example from the study application showing (1) supplemental markup and value masking, (2) interactive clickable interface, (3) feedback privacy meter, and (4) the privacy budget (solid red line on the meter). The visual markup highlights discrepancies, provides information about name frequency, and hides common values. The box on the last row indicates that the user has moused over this area, and is considering whether to click open or not. The user should be taking into account the feedback meter on top which indicates the accumulative disclosures to now in blue, and what additional risk will occur if the selected information is clicked open in orange. Finally, the solid red line on the meter indicates a limit to the disclosure that the user can request.



[Kum2019][KRIRLS19] This paper presents an interactive interface where the user starts with fully masked de-identified data and later clicks to open when more information is required for making better decisions. This is a system that reduces privacy risk through on-demand incremental information disclosure. Violin plots have been used to analyze the test results and test statistical differences in measure like error rate, k- Anonymity Privacy Risk (KAPR) scores and duration.

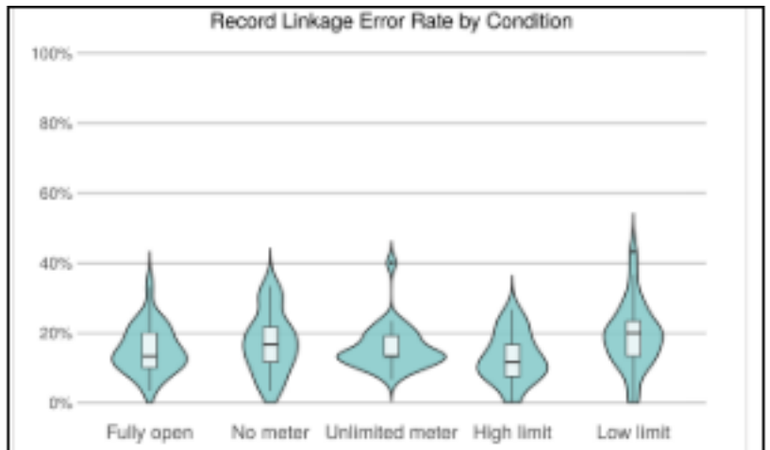


Figure 6: Percent of incorrectly linked pairs from the five conditions. Lower values indicate better performance.

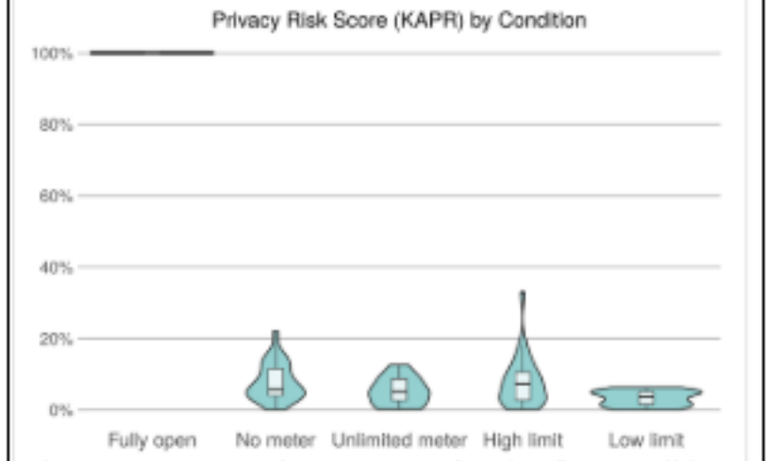


Figure 7: KAPR privacy scores for the five conditions. Lower scores indicate lower risk. Note the *fully open* condition has 100% privacy risk score due to all characters being visible by default.

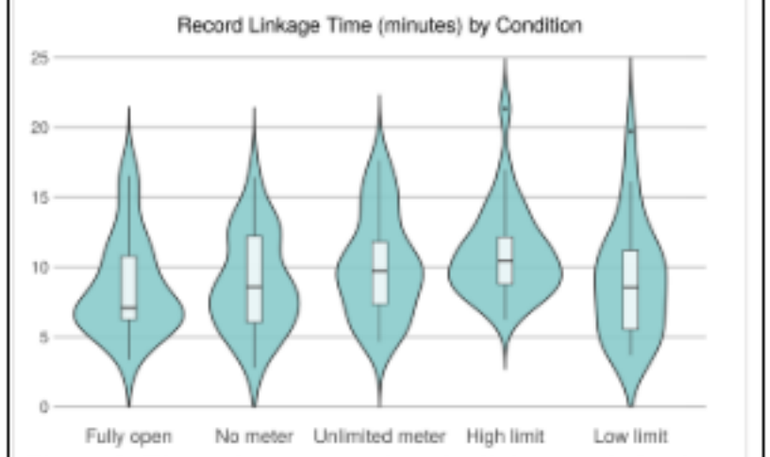
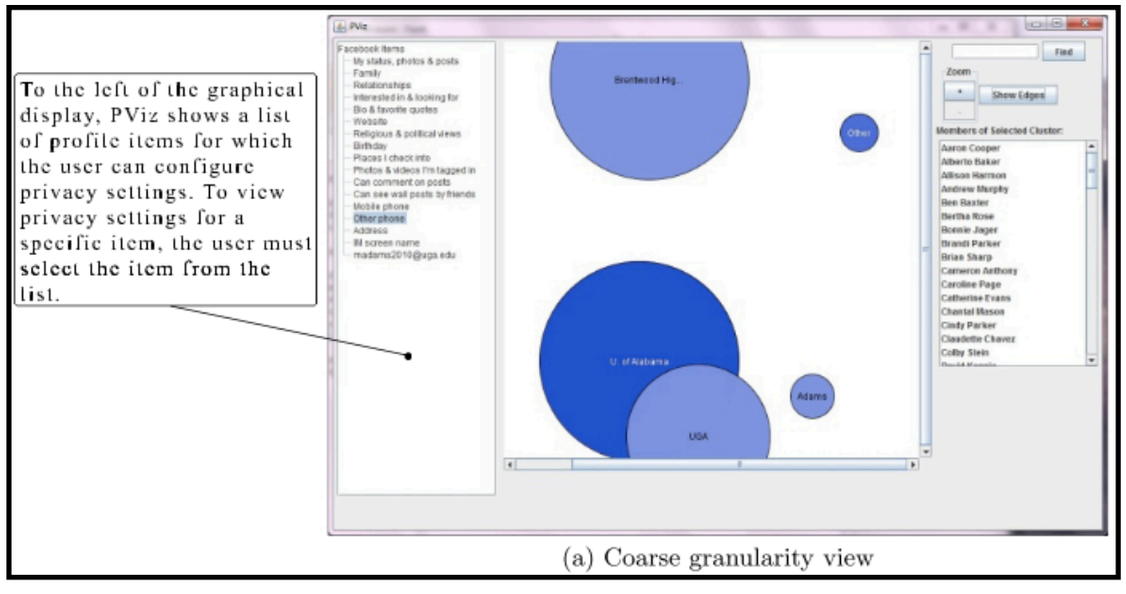
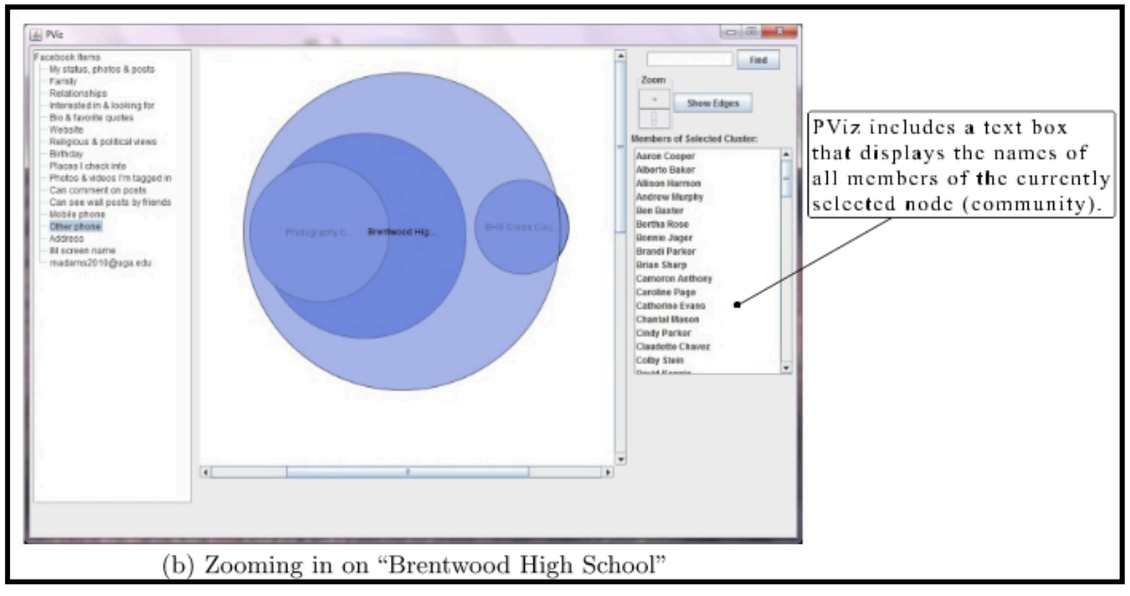


Figure 8: Time taken to complete the linkage task for the five conditions.

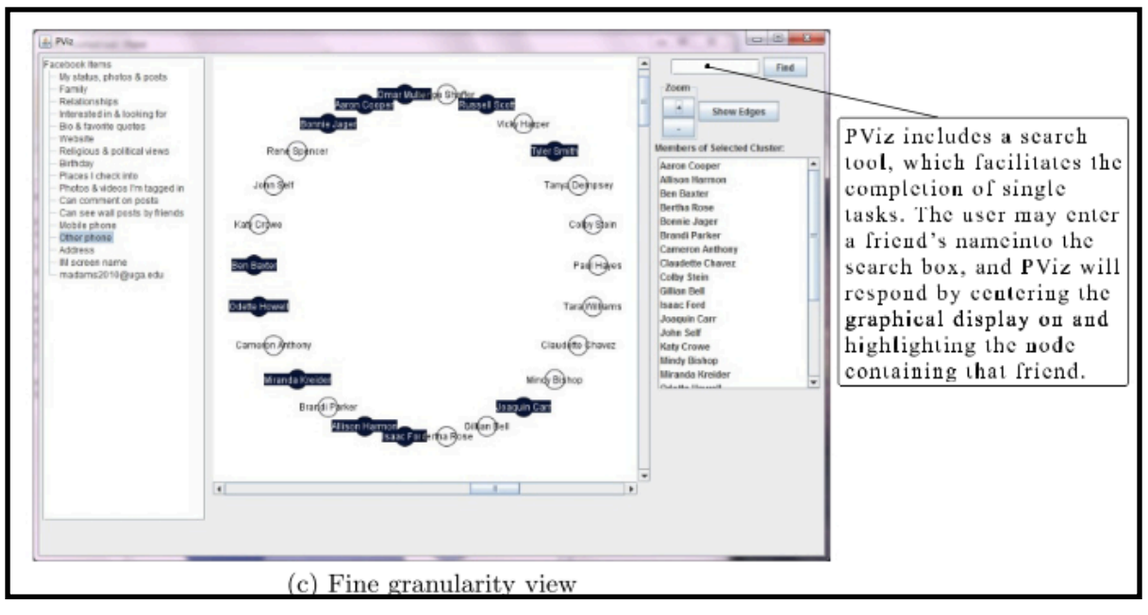
[Kum2019][KRIRLS19]



(a) Coarse granularity view



(b) Zooming in on "Brentwood High School"



(c) Fine granularity view

[Mazzia2012][MLA12] This paper presents an interface, PViz, to understand how users model groups and privacy policies apply to the existing social networks like Facebook and Google+. They state that as the existing tools present solutions to the policy comprehension problem for individuals (single tasks), they fail to scale to groups. Hence a visualization interface is required that provide the user with complete information – information about both group membership and about the privacy policies applied to these groups. The tool PViz also lets the user to comprehend privacy settings at different levels of granularity. Gradient encoding has been performed for visibility, which ranges from from 0% visibility (light) to 100% visibility (dark) and is assigned based on the user's privacy selection for a selected profile item.

Fig. 6: Rank 1 Cluster by MCODE (include loops = false, degree cutoff = 2, haircut = true, fluff = false, node score cutoff = 0.2, k-core = 2, and max. depth = 100)

[Takano2014][TOTAI14] This paper proposes a visualization system to create awareness for web activity tracking. Web tracking can be a serious privacy concern as this may reveal important user attributes. This paper has developed a web tracker identification methodology using methods like HTTP Referrer graph Analysis, Domain Aggregation, DNS-SOA-Record-Based-Grouping and Weighted Site Ranking of user data leakage. The interface, MindYourPrivacy was tested on around 129 attendees of a camp and it was inferred that the visualization of a user's network traffic should contain more relevant information by analyzing cookies which are used to identify and track users.

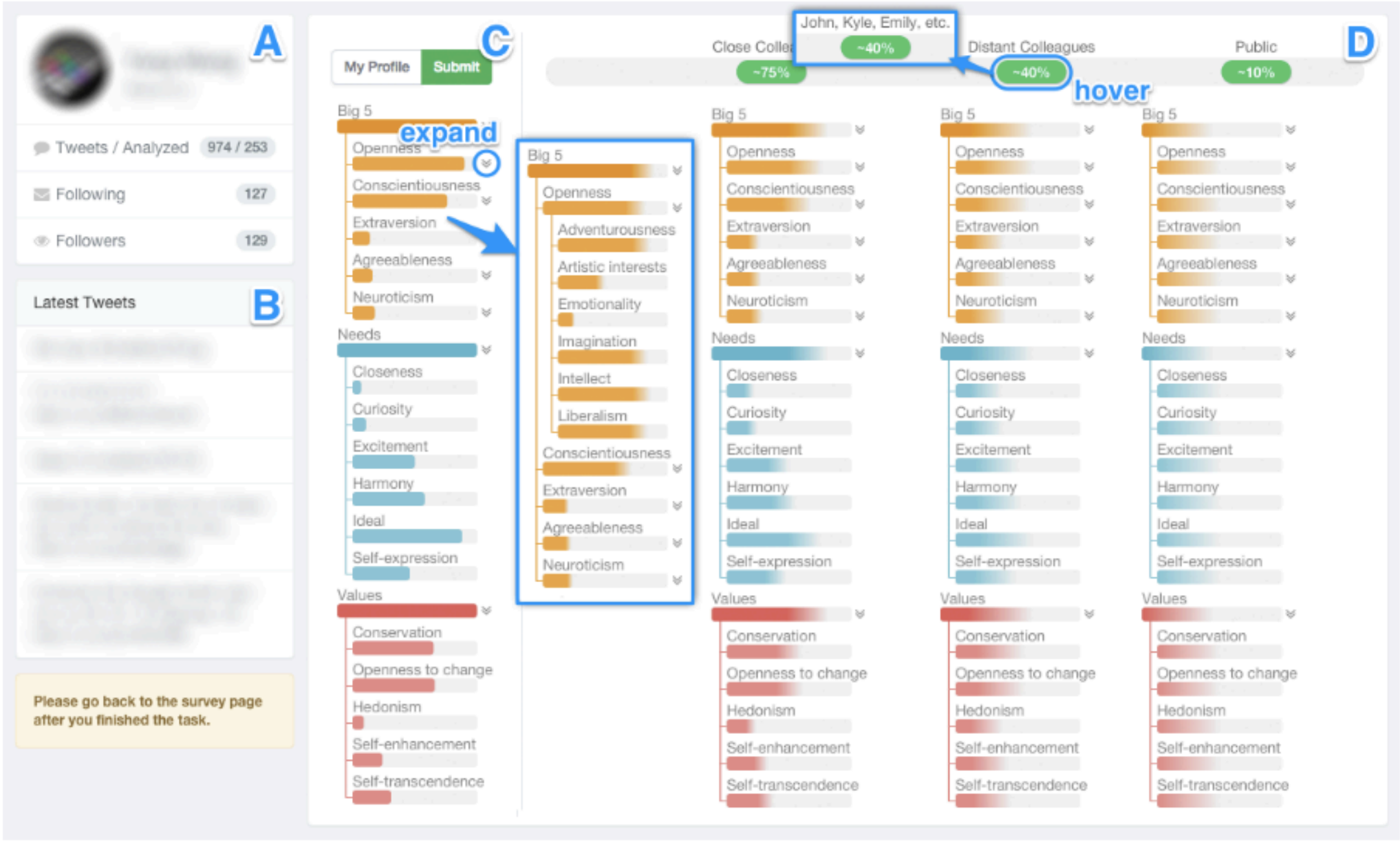


Figure 1: The screenshot of the VeilMe interface. Panel A & B: user's twitter profile and the latest tweets; C: portrait exploration panel; D: privacy setting panel. User can click to expand to reveal traits with sub-traits. When hovering a social distance knob, the input audience names of that group will be shown for user engagement. (Refer to the PDF document for better image quality.)

[Wang2015][WGXZYB15] This paper talks about a visualization tool named VeilMe which gives users the power to change their privacy settings while generating the personality profile, derived from social media. Since the personality profile could be used at workplaces, it is quite difficult for users to comprehend it and set the correct privacy policy for each trait. This novel visualization tool gives the user an easy and intuitive interface to control the privacy settings based on several parameters. One of these important parameters is "social distance", i.e., the social remoteness of the target audience from the user.

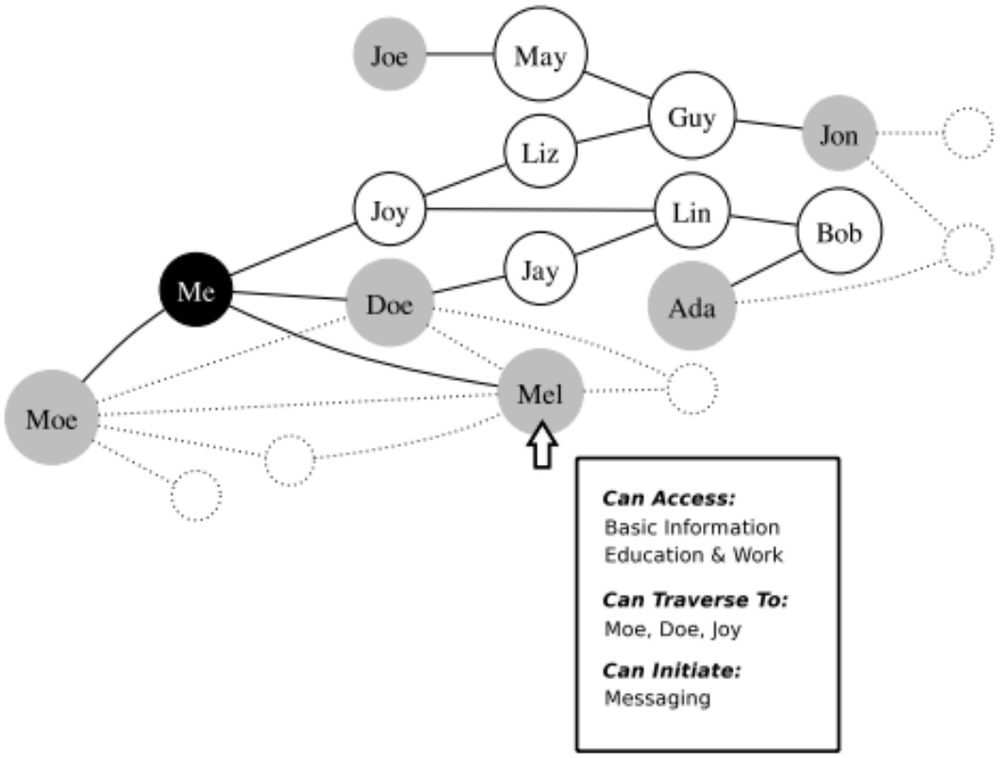


Figure 2: Visualization for reflective policy assessment.

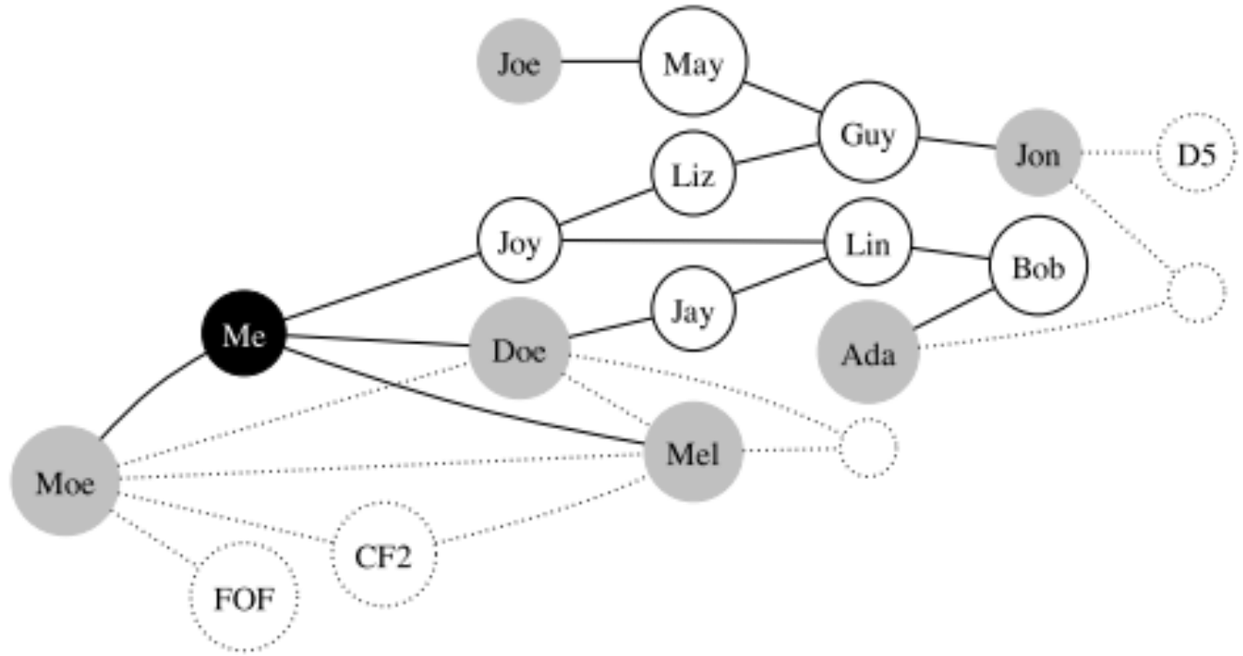


Figure 3: The extended neighborhood of a profile owner.

[Anwar2009][AFYH09] This paper is based on the hypothesis that the proper visualization of an individual's extended neighborhood in an Online Social Network (OSN) may help the individual understand the privacy implications of their access control policies. But this may compromise the privacy of others if an individual has unrestricted view of their neighborhood. Thus this paper proposed a privacy-sensitive visualization interface consisting of social graphs which will help to conduct policy assessment in a fair and meaningful manner.

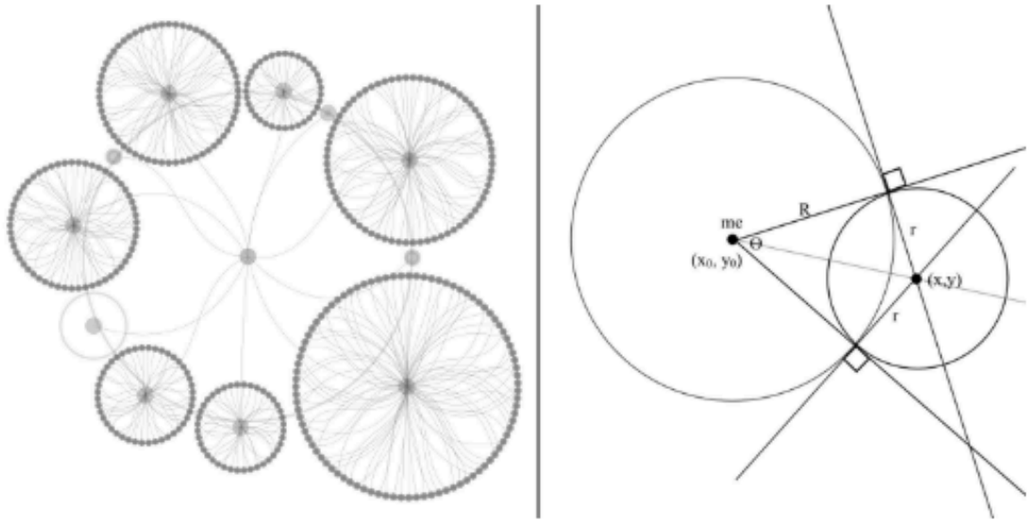


Fig. 2. Left: an overview of the circles' layout. Right: an illustration of drawing a circle around the ego.

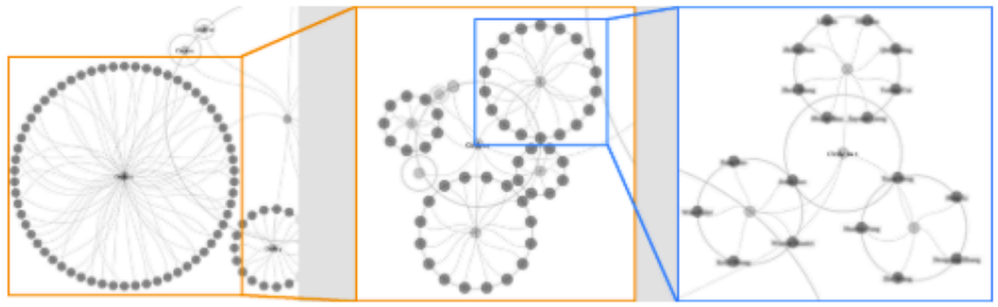


Fig. 3. An illustration of the hierarchical circles driven by a user's zooming. The names of the alters are blurred in this example to protect the user's privacy.

[Gao2013][GB13] This paper has applied community discovery methods to automatically generate friend circles for a user in Online Social Networks (OSN). This paper also implemented a visualization interface (hierarchical circles) to help the users check the visibility of their online posts and protect their privacy. These hierarchical circles have some fine-grained control settings and thus, helps the user to interact with these circles and make decisions for the visibility of their posts.

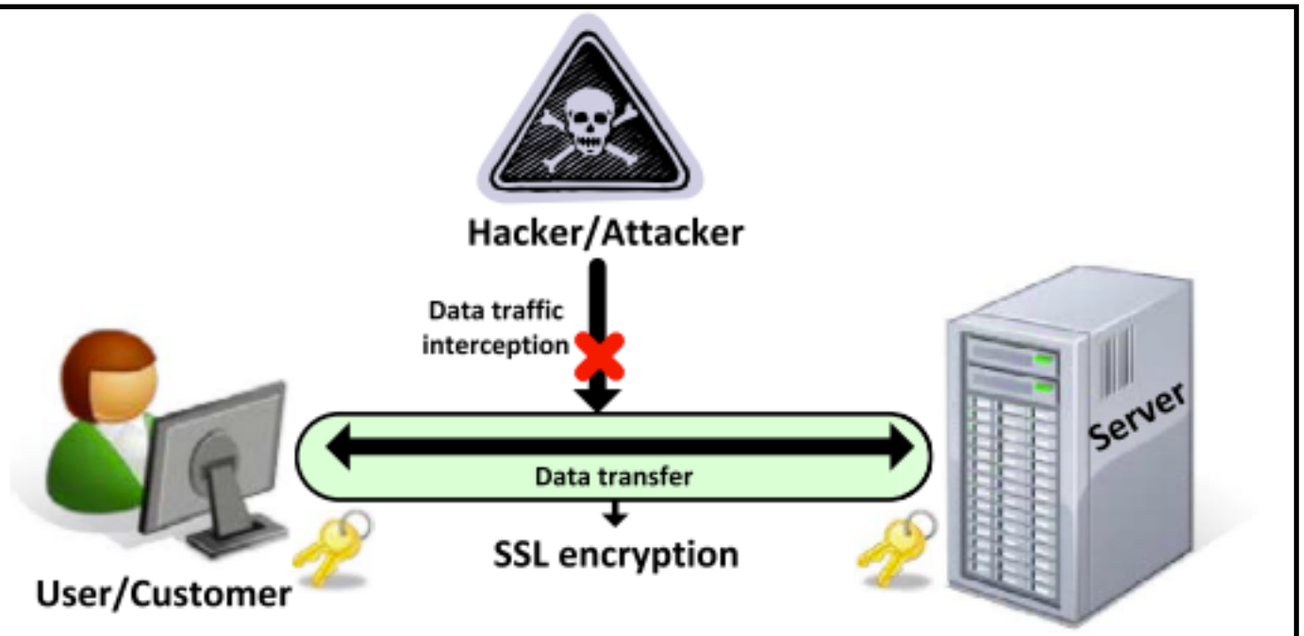


Figure 2. Visualization of SSL encryption

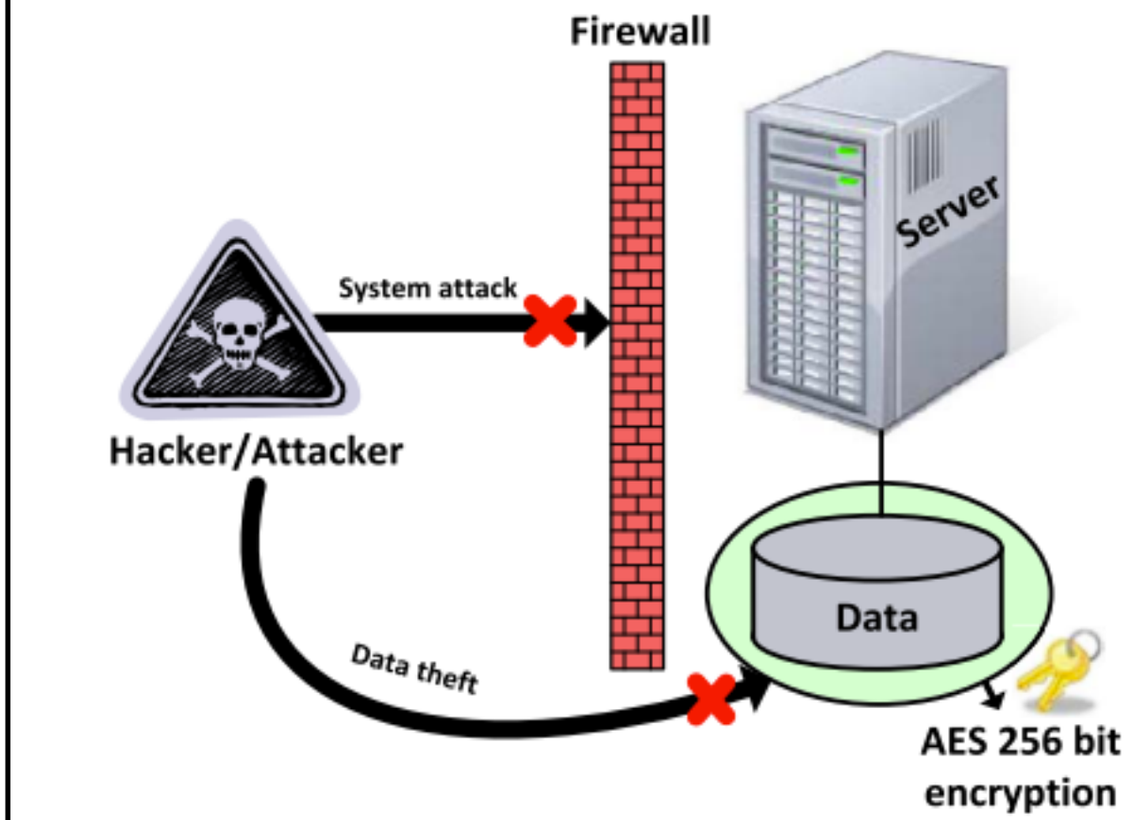


Figure 3. Visualization of a firewall and AES data encryption

[Becker2014][BHOK14] This paper reflects whether using visualizations to communicate privacy and security measures have positive effects on trust. The paper used certain infographics to depict certain privacy concepts like SSL encryption, AES encryption and studied the improvement on privacy and trust. The study concluded that though these descriptive images have a positive effect on the trust in the provider, there was no significant improvement regarding data security and privacy, in comparison to the text-based privacy policy.

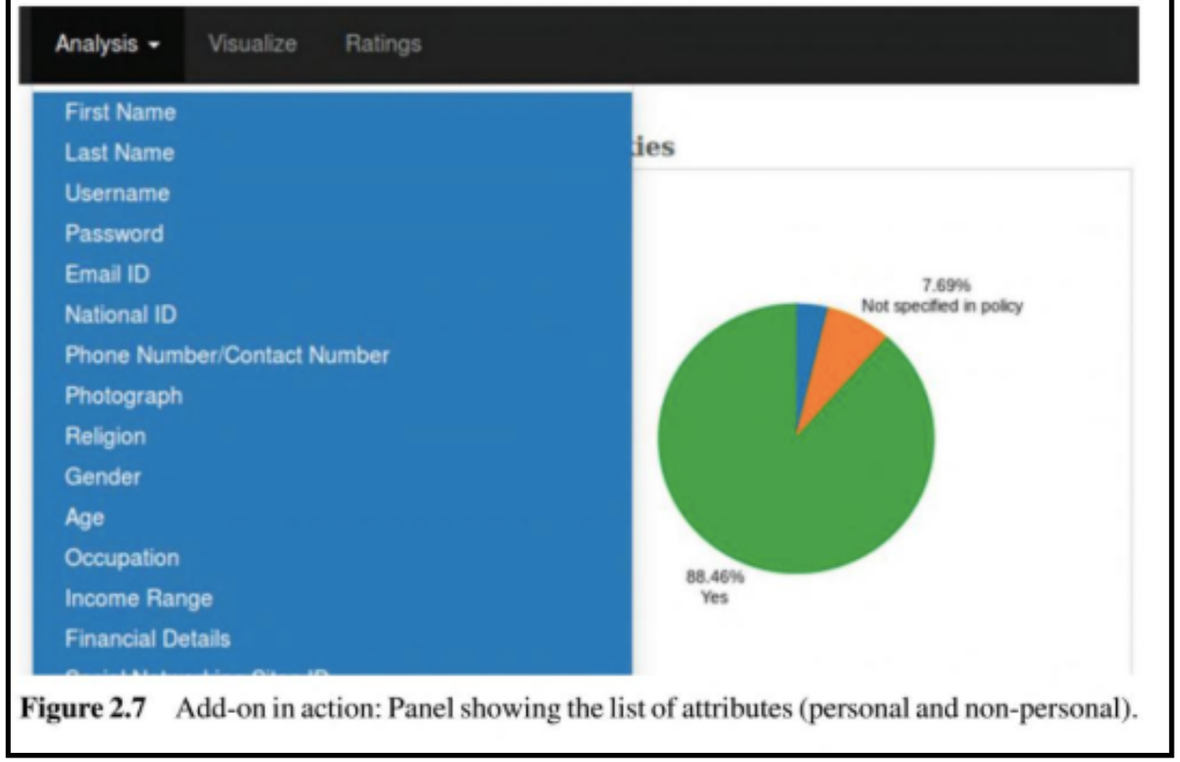


Figure 2.7 Add-on in action: Panel showing the list of attributes (personal and non-personal).

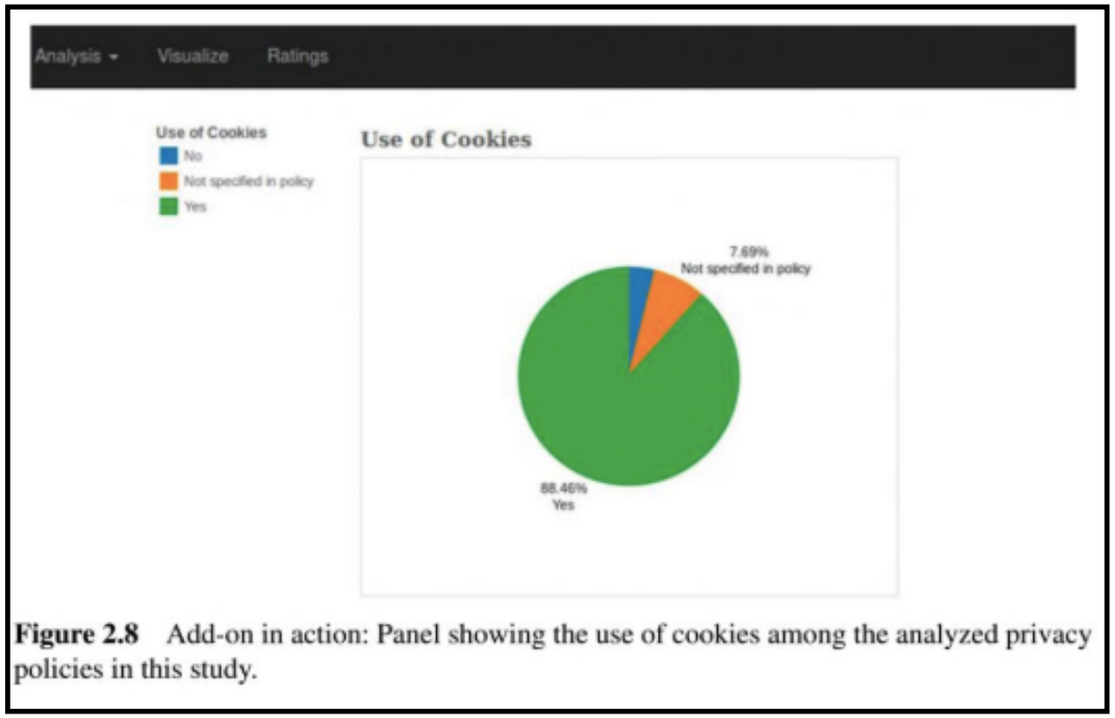


Figure 2.8 Add-on in action: Panel showing the use of cookies among the analyzed privacy policies in this study.

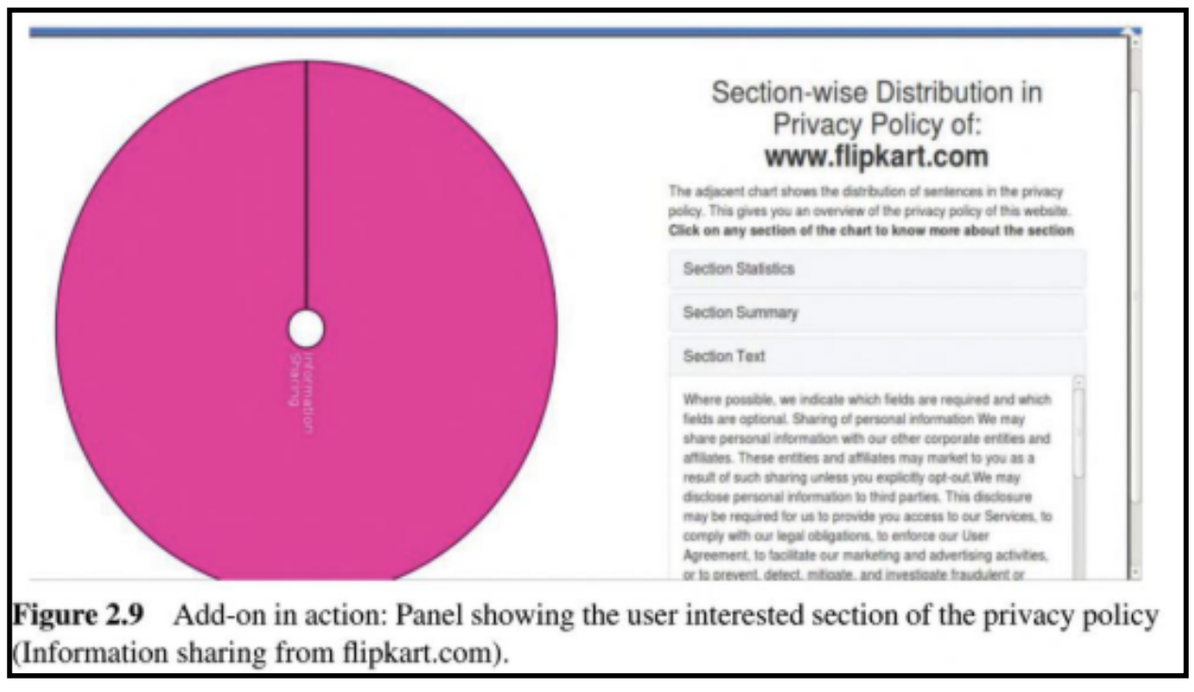


Figure 2.9 Add-on in action: Panel showing the user interested section of the privacy policy (Information sharing from flipkart.com).

[Dhotre2017][DBKO17] This paper has implemented a method to perform semi-automatic analysis of the privacy policies of certain websites and generate visualization in order to help the user understand the policies better. This visualization interface, consisting of pie charts, helps the user understand the use of different Personally Identifiable Information (PII) by the website, according to their privacy policies. The interface also summaries certain sections like use of cookies, information sharing polices and help the users to understand them better. The Privacy Policy Elucidator Tool (PPET) collects the privacy policies from different websites, parse them, classify them using machine learning techniques like Naïve Bayes classifier and uses the extracted paragraph and summary for the visualization. It also evaluates the trustworthiness of the website and displays the same through a donut visualization.

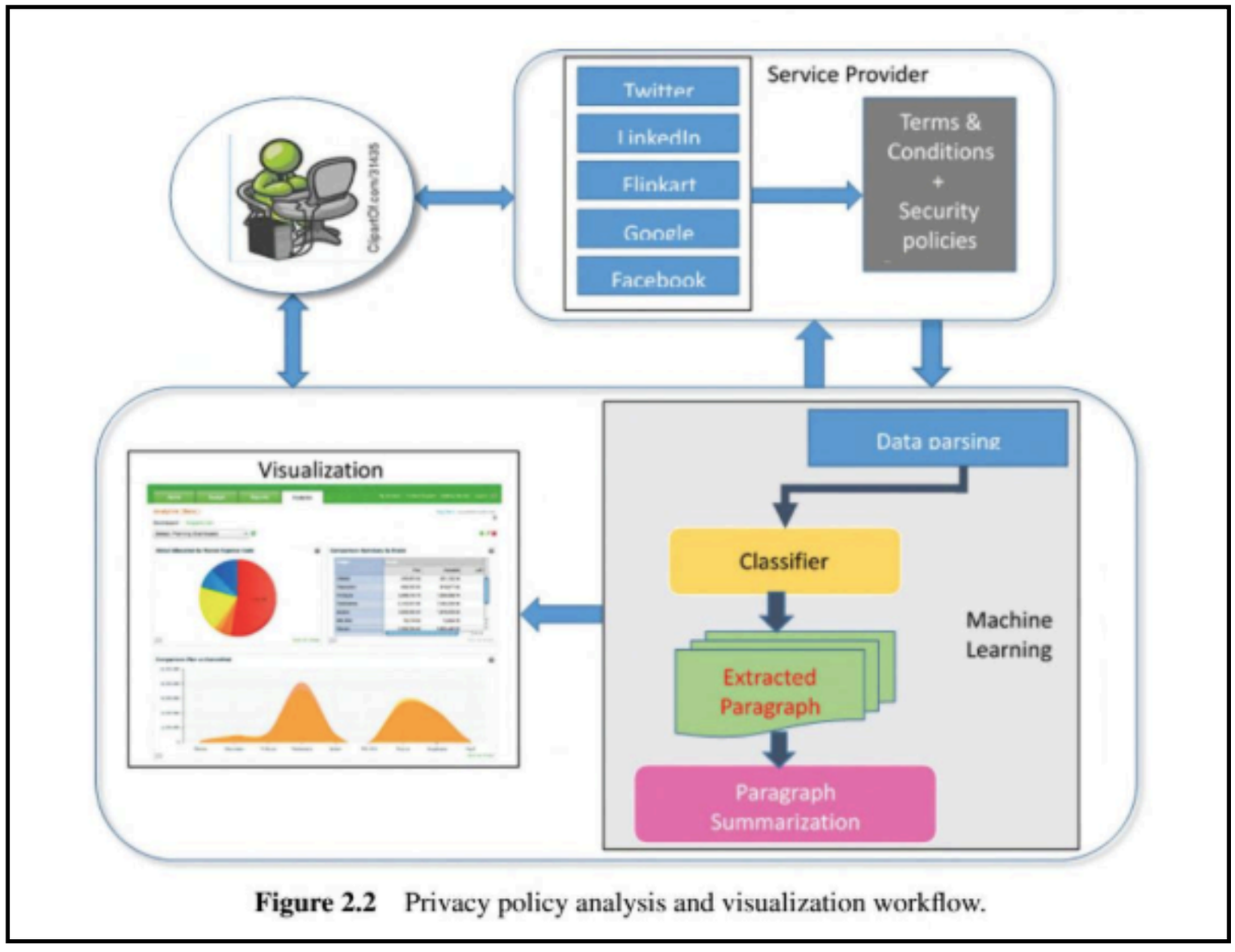


Figure 2.2 Privacy policy analysis and visualization workflow.

[Dhotre2017][DBKO17]

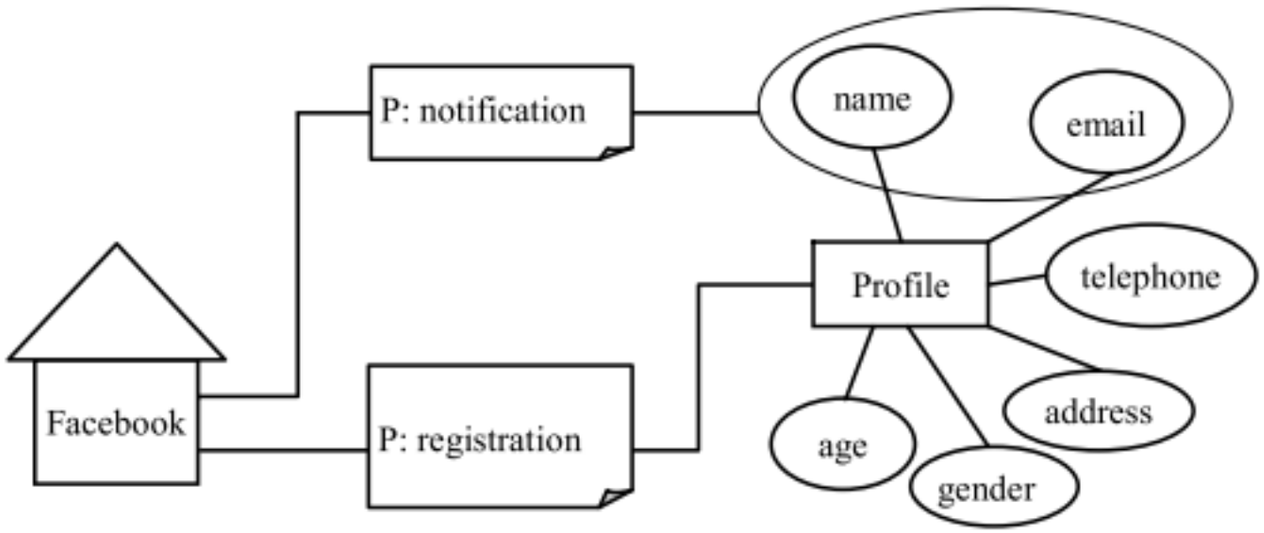


Figure 10. PPVM of Example 2

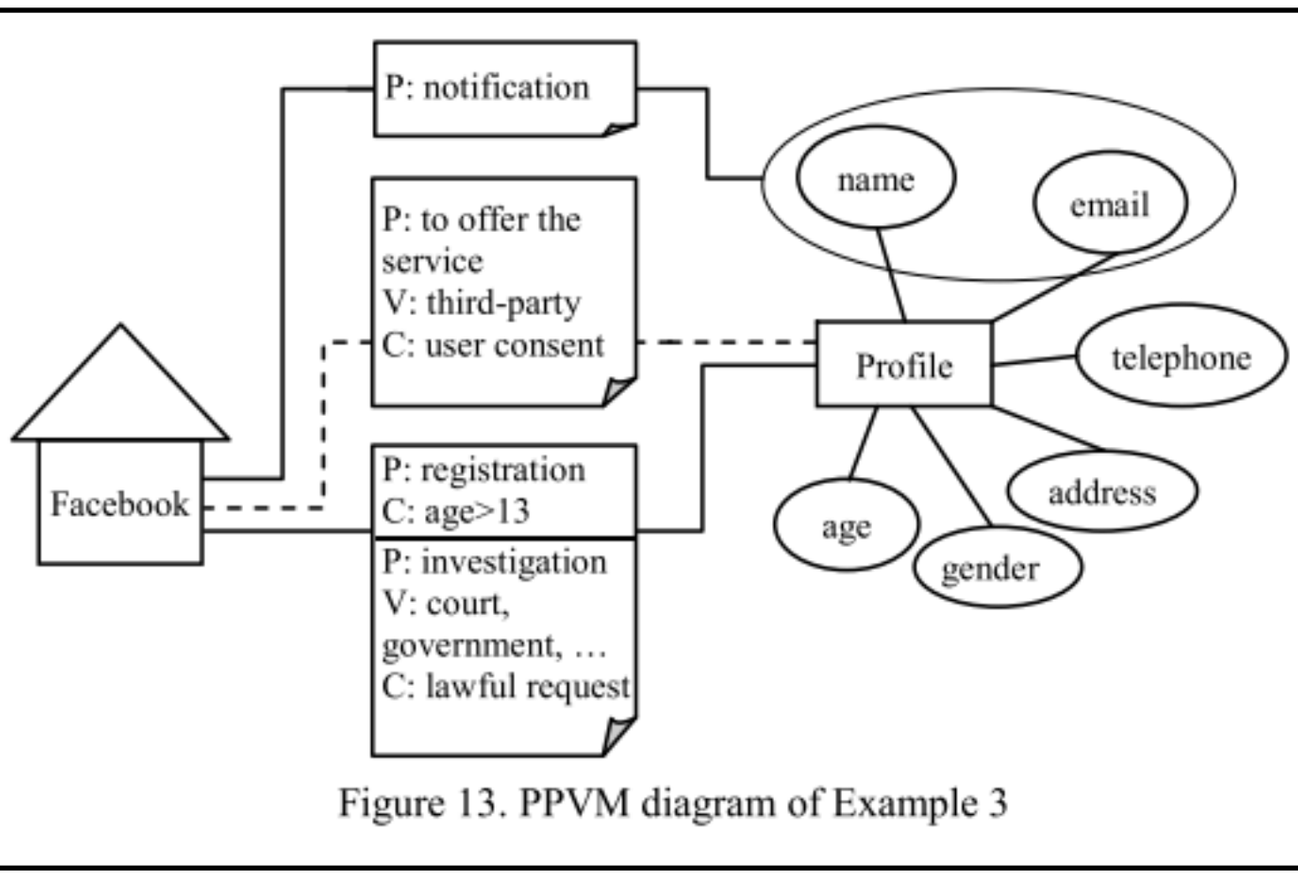


Figure 13. PPVM diagram of Example 3

[Ghazinour2009][GMB09] This papers presents a visualization model which will help the data owners understand the privacy policy of a website and help the policy officers to better understand the designed policies. The figures present the Privacy Policy Visualization Model (PPVM) for certain online social network websites. These relationship diagrams help understand privacy policies of these websites like using the users's name and email address to send notifications regarding new services, not collecting data of anyone under a certain age limit, disclose user information pursuant to lawful requests etc. The model suggests to highlight the purpose(P), granularity(G), visibility(V), retention(R) and constraint(C) of the privacy policies in this relationship diagrams.

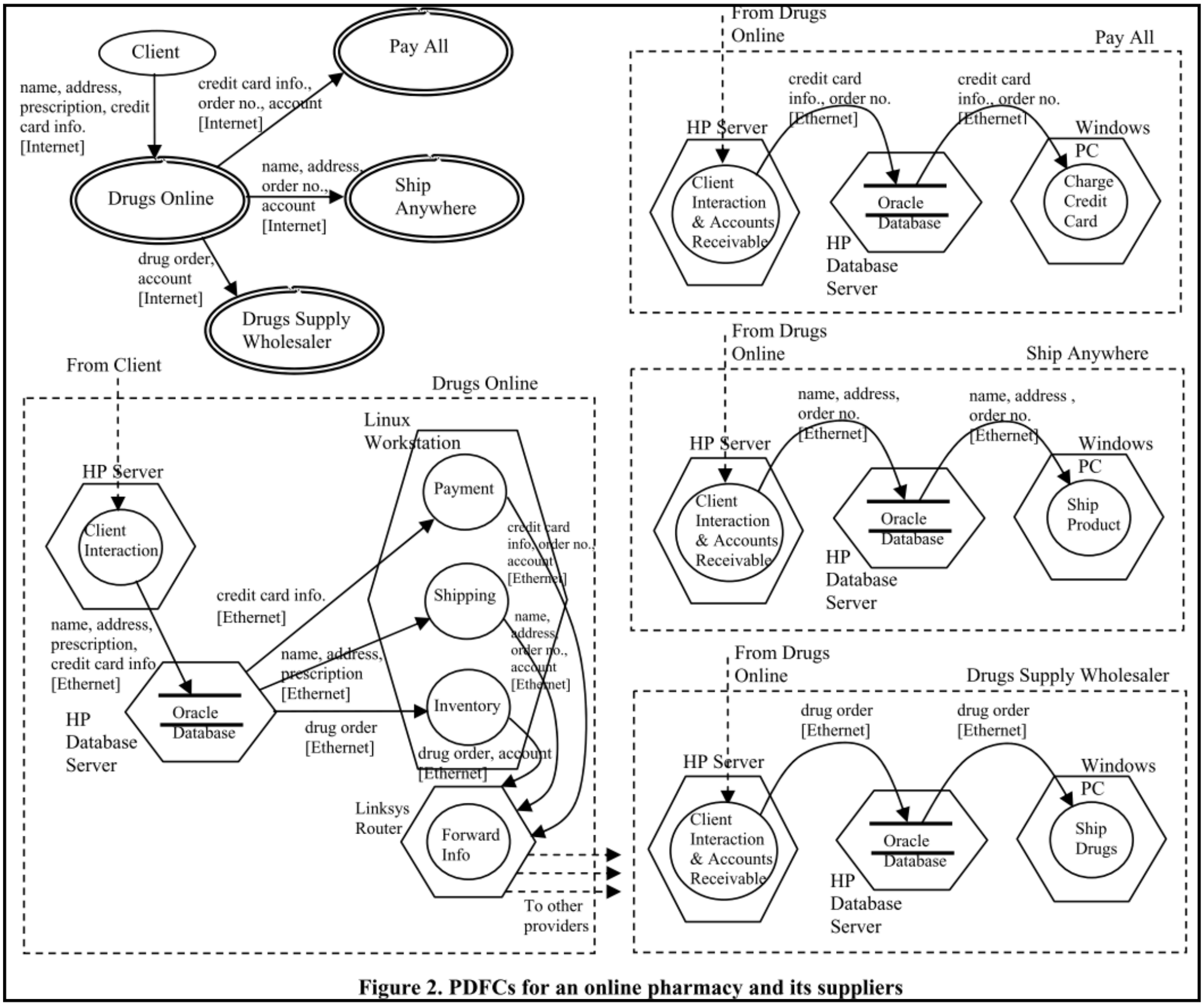


Figure 2. PDFCs for an online pharmacy and its suppliers

[Yee2006][Y06] This paper presents a visualization interface which will help the privacy and security analysts understand the flow of information between different organizations and identify vulnerabilities due to policy non-compliance. The figure explains how information flows between an online pharmacy and it suppliers.

No Visualization

[Hongde2014][HSH14] This paper explores the realm of differential privacy in a cloud-based research environment. It has dome certain improvements on the k-means clustering algorithm by combining differential privacy along with k-means clustering. This research will address the challenges faced with location-based information, specially the privacy concerns. This paper also proposes to perform the aggregation operation on the large-scale sample data in order to protect the privacy of the data.

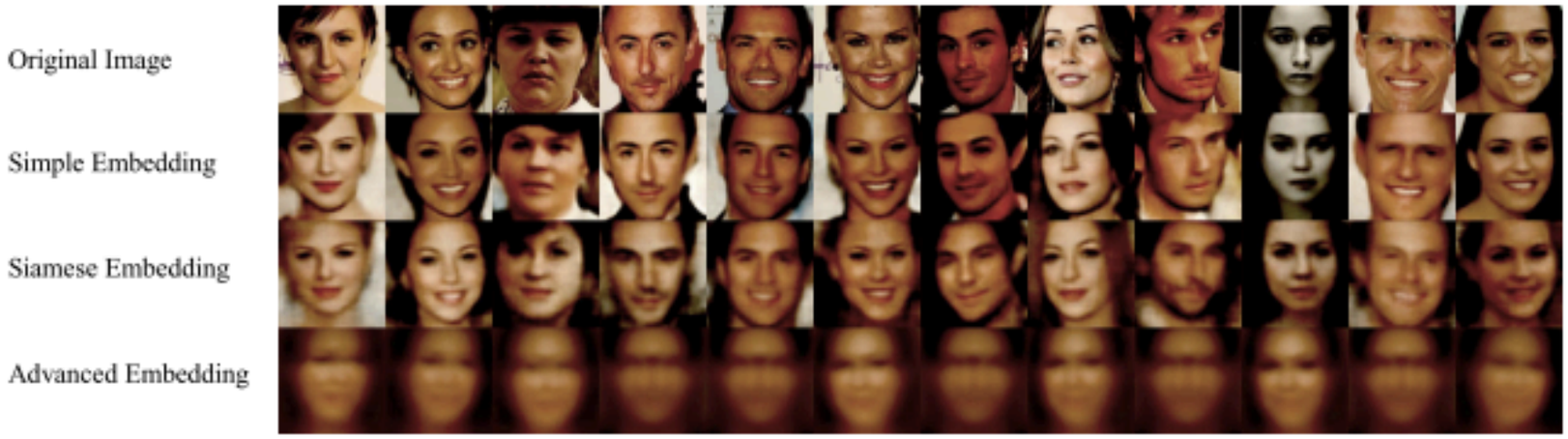


Fig. 11: The first row shows the original images and the others show the reconstructed ones from exclusive feature. In all reconstructed images, the gender of the individuals is recognized to be the same as the originals. In addition, From simple to advanced embedding, the identity of the individuals is increasingly removed, proving that the *advanced embedding* has the best privacy preservation performance.

[Osia2020][OSSTKRKH20] This paper address the privacy and efficiency challenges faced while performing continuous user data collection in order to feed the machine learning models. Dimensionality reduction, Siamese fine-tuning and noise addition have been suggested as some of the methods to protect privacy in this hybrid user-cloud framework proposed by this paper. This paper also suggests the use of an Auto-Encoder visualization technique which will help in obtaining insights into the privacy of images.

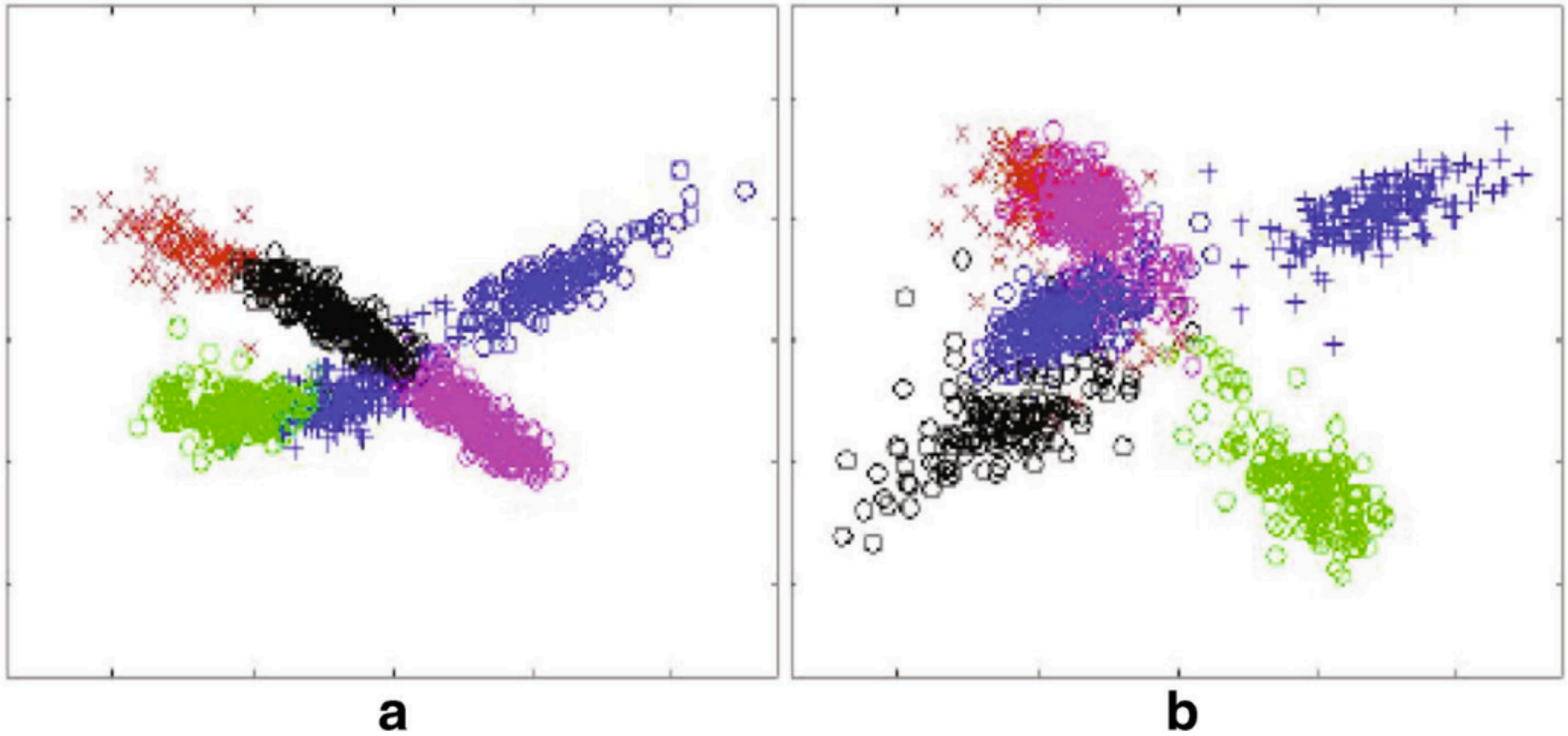


Fig. 7 Visual comparison of the results of two iterative clustering methods. **a** Torre and Kanade's Discriminant Clustering Analysis and **b** An iterative DCA approach: version "$\frac{n}{\ln(n)}$-Decrease" (courtesy from Xuyang Lu 2015)

[Kung2017][K17] This paper has important contribution towards the privacy protection of big data. Dimensionality reduction is an important method to comprehend data of large scale and to protect privacy of the users mentioned in the dataset. Hence, this paper uses Discriminant Component Analysis (DCA), a supervised version of Principal Component Analysis (PCA) for the visualization because DCA can support data of high compression (small dimensionality) and the recoverability can be controlled. This paper has also compared among the results of different clustering methods like Torre and Kanade's DCA and an iterative DCA approach.

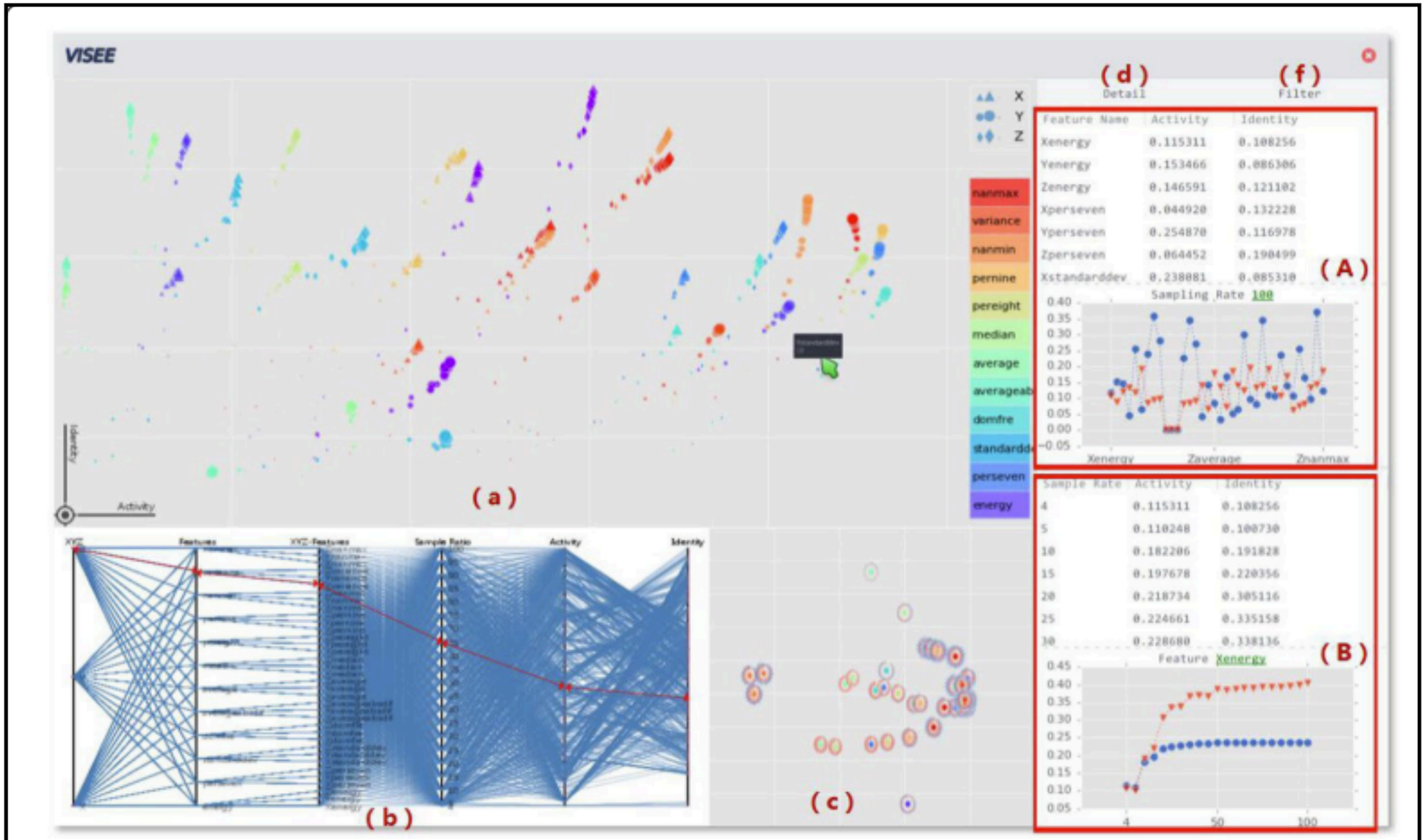


Fig. 10 The interface of the visualization tool, VISEE

[Xiao2018][XLZMMXLL18] This paper presents a visualization tool named VISEE which will help to maintain the balance between high application utility and less privacy leakage in the case of sharing of sensor data. Accelerometer data collected from different mobile devices has been used as an example. The visualization containing parallel coordinates, feature grid diagram, ranking chart will help to select the appropriate combination of features and sampling rates, thus making a good decision on the trade-off between utility and privacy.