




# Automated Refined Comic Generation: From Investigation Provenance to Data Comics using Visual Narrative Structure

K. A. Roggenbuck<sup>1</sup> , A. Vilanova<sup>1</sup> , and S. van den Elzen<sup>1</sup> 

<sup>1</sup>TU Eindhoven, The Netherlands

## Abstract

Visual analytics has become an important approach for criminal investigations due to the increasing amount of physical and digital data related to cases. Although state-of-the-art tools are used daily to search for evidence in the data and report the investigator's findings, building such reports remains a labor-intensive manual process. Furthermore, these reports commonly contain only a manually selected set of the investigation results, but not how these results were derived. This lack of information about the chain of evidence not only weakens reproducibility and transparency, but also makes the evidence vulnerable by jurists in court. Instead of textual reports we believe annotated visuals of the actual data exploration process better portray what the investigators did and how they came to the evidence. To this end, we introduce ARC, a framework for automatically generating comic summaries for digital investigations based on the Visual Narrative Structure from comic theory. Especially, ARC is the first framework that fully automatically generates and refines comic summaries based on interactions with investigation tools.

## CCS Concepts

• **Human-centered computing** → *Visual analytics; Visualization techniques;*

## 1. Introduction

Digital forensics plays a key role in today's law enforcement process. To efficiently utilize digital evidence, (visual) investigation tools such as Analyst's Notebook [i2A24] or Hansken [han24] are used to process data and to build and test hypotheses. Once evidence is found, it is described in a report in written form with the possible addition of screenshots of the tool. These screenshots are usually manually recorded and must be processed and selected by the investigator to integrate them into a report. With this process, the provenance within investigation tools becomes an optional addition to the work of investigators, leaving them the choice of what to include and what to omit. Furthermore, this process is labor-intensive and adds additional burden to the investigators' daily work while they are already under constant time pressure. Therefore, we believe that an automated creation of visuals that represent the investigation process could help addressing these problems.

While multiple approaches are promising for this, recent work has shown that when combining text and visualizations – as with descriptions and screenshots in reports – Data Comics [ZME15] are in certain scenarios more understandable than other approaches, such as illustrated texts or infographics [WWF\*19]. However, state-of-the-art approaches, such as Graph Comics [BKH\*16], ProvenanceComics [SS17] or DG Comics [KLN\*24], require significant user effort for either every comic creation or refinement, making these approaches impractical for investigators daily work. Nevertheless, previous work also studied the structure of

comics in general in the form of the Visual Narrative Structure (VNS) [Coh13b], which opens the possibility for structured analysis and refinement of (data) comics. In this work we propose the **A**utomated **R**efined **C**omic (ARC) framework which applies the VNS to Data Comics for automatically generating comic summaries for investigations.

## 2. Visual Narrative Structure

In the Visual Narrative Structure (VNS) [Coh13b], comic panels are classified into the following Basic Narrative Categories (BNC):

- Establisher (E): Setup a scene without meaningful action,
- Initial (I): Initiates the action of the sequence,
- Prolongation (L): Delays the Peak of the sequence,
- Peak (P): The height point of the narrative's tension, and
- Release (R): Serves as a follow-up to the Peak.

Based on these categories, a general sequence in a comic can be described in the form

$$(E) \rightarrow (I(L)) \rightarrow P \rightarrow (R), \quad (1)$$

where brackets denote optional categories, making the Peak the only category that must exist in a sequence. Furthermore, based on their role in the sequence, the following priority of panels exist:

$$Peak > Initial > Release > Establisher > Prolongation. \quad (2)$$

While all roles can be important to make a sequence funny or exciting, these priorities indicate the panel’s importance for *understanding* the sequence. Furthermore, to increase scope and flexibility of the VNS, *combinatorial structures* are introduced to build a hierarchy resembling a tree structure. For example, multiple consecutive panels with the same BNC can be abstracted to one instance with the same class for the whole sequence.

### 3. Automated Refined Comics

To utilize the **A**utomated **R**efined **C**omic (ARC) framework for the automatic creation and refinement of comic summaries for criminal investigations, a domain expert needs to set it up once per tool. This setup involves the assignment of BNCs to actions within the investigation tool. Afterward, ARC can automatically be run on future investigations utilizing its *preparation* and *refinement* phase.

The input to ARC’s *preparation phase* are captured states from the investigation tool, for example by using CLUE [GLG\*16]. Based on these, the preparation phase derives the BNC per state, builds its hierarchic structure, assigns priorities to the states, and splits the whole investigation in sub-stories. While most of the performed steps are unique for ARC, splitting the story into sub-stories was already proposed by ChartStory [ZXC\*21]. However, ChartStory’s approach is suitable only for charts, while ARC’s is generalized for arbitrary tracked states within investigation tools.

Subsequently, ARC performs its *refinement phase* based on the preparations made. These refinements include merging states, highlighting content, adding annotations, and calculating a suitable layout. While these steps separately are all addressed in previous literature, we present major differences in the adding of annotations and the calculation of layouts. Compared to DG Comics [KLN\*24] head bar annotations, ARC automatically places speech bubbles pointing to the action which eases the users understanding of the panel. Furthermore, in contrast to the layout calculations from ChartStory [ZXC\*21] and Calic et al. [CGC07], ARC finds for the fixed order of states an optimal z-layout – read from left to right and row by row – for increased reader understanding [Coh13a]. An example comic summary from a proof-of-concept implementation is shown in Figure 1.

To assess ARC’s suitability for communicating investigation procedures to investigators and jurists, we conducted online surveys and think-aloud sessions with in total eleven investigators and five jurists. Based on comprehension questions for four comic summaries from three investigation tools, we found that more than 80% were answered correctly across the different use cases and participant groups. While this is only an initial evaluation, it shows that ARC has the potential to make a positive impact on investigations by summarizing its process.

### 4. Future Work

The automatic generation of refined comic summaries opens possibilities in different directions. The next step would be to integrate ARC in the daily work of investigators and study its implications on their work. While ARC presents a first step for reporting the process of finding digital evidence, field studies can help to further improve the framework for specific real world use cases.

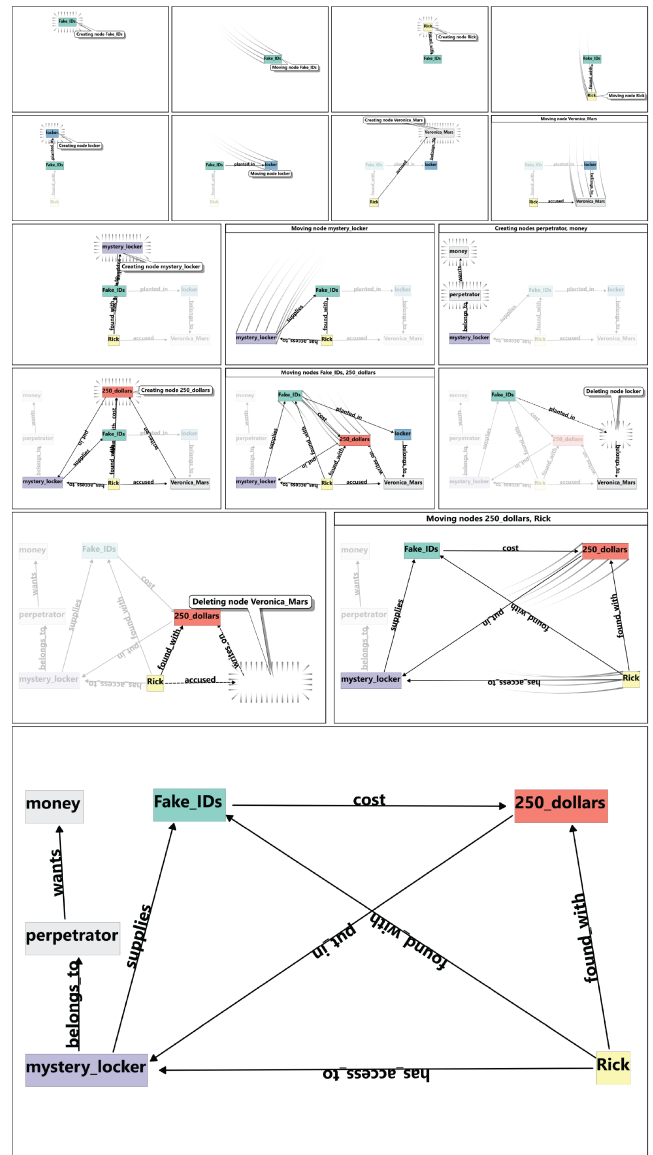


Figure 1: An example comic summary generated with ARC where the panels include refined representations of investigation states.

Furthermore, while we demonstrate ARC for criminal investigations, we believe that it could have applications beyond this field. For example, recent work has discovered the benefits of comics in school education [Akc20] as an engaging medium for complex topics. In this area, teachers could benefit from ARC, for example, by using programs like Microsoft Excel or Power Point and generating comic summaries for students to ease understanding.

### Acknowledgements

This work was funded by NWO through grant KICH1.VE01.20.011 (“AI4Intelligence”). Furthermore, we thank the participants of our evaluation for their time and productive comments.

## References

- [Akc20] AKCANCA N.: An alternative teaching tool in science education: Educational comics. *International Online Journal of Education and Teaching* 7, 4 (2020), 1550–1570. [2](#)
- [BKH\*16] BACH B., KERRACHER N., HALL K. W., CARPENDALE S., KENNEDY J., HENRY RICHE N.: Telling stories about dynamic networks with graph comics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 3670–3682. [1](#)
- [CGC07] CALIC J., GIBSON D. P., CAMPBELL N. W.: Efficient layout of comic-like video summaries. *IEEE Transactions on Circuits and Systems for Video Technology* 17, 7 (2007), 931–936. [2](#)
- [Coh13a] COHN N.: Navigating comics: An empirical and theoretical approach to strategies of reading comic page layouts. *Frontiers in psychology* 4 (2013), 186. [2](#)
- [Coh13b] COHN N.: Visual narrative structure. *Cognitive science* 37, 3 (2013), 413–452. [1](#)
- [GLG\*16] GRATZL S., LEX A., GEHLENBORG N., COSGROVE N., STREIT M.: From visual exploration to storytelling and back again. In *Computer Graphics Forum* (2016), vol. 35, Wiley Online Library, pp. 491–500. [2](#)
- [han24] Hansken the open digital forensic platform. <https://www.hansken.nl/>, 2024. Accessed: 2024-07-04. URL: <https://www.hansken.nl/>. [1](#)
- [i2A24] i2 analyst's notebook. <https://i2group.com/i2-analysts-notebook>, 2024. Accessed: 2024-03-28. URL: <https://i2group.com/i2-analysts-notebook>. [1](#)
- [KLN\*24] KIM J., LEE H., NGUYEN D. M., SHIN M., KWON B. C., KO S., ELMQVIST N.: Dg comics: Semi-automatically authoring graph comics for dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics* (2024). [1](#), [2](#)
- [SS17] SCHREIBER A., STRUMINSKI R.: Visualizing provenance using comics. In *9th USENIX Workshop on the Theory and Practice of Provenance (TaPP 2017)* (2017). [1](#)
- [WWF\*19] WANG Z., WANG S., FARINELLA M., MURRAY-RUST D., HENRY RICHE N., BACH B.: Comparing effectiveness and engagement of data comics and infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–12. [1](#)
- [ZME15] ZHAO Z., MARR R., ELMQVIST N.: Data comics: Sequential art for data-driven storytelling. *tech. report* (2015). [1](#)
- [ZXC\*21] ZHAO J., XU S., CHANDRASEGARAN S., BRYAN C., DU F., MISHRA A., QIAN X., LI Y., MA K.-L.: Chartstory: Automated partitioning, layout, and captioning of charts into comic-style narratives. *IEEE Transactions on Visualization and Computer Graphics* 29, 2 (2021), 1384–1399. [2](#)