Approaching Asian Popular Culture Beyond Japan from a Metadata Analytics Perspective:
Views Opened up by the TINTIN and JVMG Projects

Zoltan Kacsuk and Neil Cohn

Mechademia Conference 2026
Traversing Trans-Asian Imaginaries: Anime, Manga, and Media Cultures
National University of Singapore, 29-30 May 2026

[1]

Hello, I am Magnus Pfeffer from Stuttgart Media University, and I will be reading this presentation in the place of my colleague Zoltan Kacsuk, who unfortunately had to cancel his trip due to unforeseen personal circumstances. This presentation is a collaborative work with Neil Cohn from Tilburg University,

[2]

and will provide a short introduction of Neil's TINTIN project and our own JVMG project, and offer two examples of how the data made available in these projects can be employed to pursue research questions about anime, manga and games beyond Japan in order to help encourage interested researchers to explore and use these data for their own inquiries.

[3]

First, a quick note on metadata analytics and the type of data we are concerned with here.

[4]

Data driven, digital and quantitative methodologies and frameworks have been gaining ground in the humanities for some time now. Metadata analytics is just one among many approaches.

[5]

It has its roots in bibliometrics and scientometrics, and emphasizes using various forms of descriptive metadata about works, characters, creators and so on, to help both corroborate previously held assumptions as well as identify new promising research questions and hypotheses based on large-scale quantitative analyses. One of the key prerequisites of this type of work is the availability of large amounts of high quality, granular metadata,

[6]

which leads us on to the two projects we would like to discuss today, both of which offer rich metadata corpora for researchers.

[7]

First up, we have Neil Cohn's TINTIN project, officially titled "Visual narratives as a window into language and cognition," which was a five-year-long European Research Council Starting Grant funded endeavor aimed at identifying cross-cultural patterns in the visual languages of comics from around the world, as well as their relationship to the languages spoken by the creators and readers of these works. The TINTIN project was the next step in the corpus building work started by Neil with his preceding Visual Language Research Corpus project. The TINTIN corpus is a truly unique achievement boasting the manual annotation by 55 annotators of 1030 comics works from 144 different countries and territories along a wide range of dimensions relevant to comics research such as panels, characters, layout, framing, backgrounds, continuity, compositional structure, emotional valence, motion events, perspective taking, gender, and so on. The full TINTIN Corpus is gradually being made available under a CC license for researchers to explore. A public interface, where people can create graphs to examine and visualize the data, is also in development.

[8]

Beyond the corpus data itself, another contribution of the TINTIN project to comics research in general is the Multimodal Annotation Software Tool or MAST for short. This tool, developed by Bruno Cardoso with Neil, provides an advanced graphical user interface enabling the precise manual annotation of comics works. MAST will be packaged in a stand-alone Docker Container and made available for the wider research community later during this year to enable the simple deployment of this tool for anyone who would like to create their own corpus of hand-annotated comics works and possibly contribute to an ever-growing family of interoperable annotated comics corpora for more granular research. In the meantime the TINTIN annotation guide is already available on the project website.

[9]

The second project, the utility of which we will be discussing today, is the Japanese Visual Media Graph or JVMG for short. This project was originally conceived by Magnus Pfeffer and Martin Roth and takes as its starting point the common experience of researchers working on topics in the domain of popular Japanese visual media that online databases created by fan or enthusiast communities are often the best source of detailed information on source materials, characters and creators. The aim of the JVMG is to connect these types of resources – in cooperation with the concerned communities where possible and always respecting the data handling guidelines and licenses of the source databases –

into a single knowledge graph to enable large-scale quantitative research on Japanese anime, manga, games and related domains. The complete knowledge graph is made available under a researcher-friendly CC attribution-non-commercial-share-alike license, and will also be made available in a Docker Container later this year enabling local deployment. This work has been made possible thanks to the generous support of the German Research Foundation.

[10]

For now the data can be accessed and browsed through the project's web frontend. This overview page showcases the most notable different data sources that have been integrated – Anime Characters Database, AnimeClick, The Visual Novel Database, Media Arts Database, and data from the Game Preservation Society – as well as their entity types and the numbers thereof. There is also a convenient search option that displays results ranked and according to data facets, as well as a custom query builder tool that provides a graphical user interface for creating complex data queries without a need for understanding the SPARQL query language or the finer details of the JVMG knowledge graph ontology.

[11]

Let us now turn towards our first example highlighting the utility of these data resources for researchers working on Asian popular culture beyond Japan.

[12]

While the TINTIN corpus does provide global coverage the works in the corpus are not evenly or proportionally distributed among countries or larger geographic regions. Fortunately for researchers interested in Southeast Asia, it is the region with the fourth highest number of works in the corpus, with 98 comics in total across all concerned countries, styles and formats. What is more, 57 of these works have been classified as manga – the umbrella term in the TINTIN corpus for the stylistic family of comics centered on Japan – with 33 of them from print formats and 24 webcomics. We have separated Japan out from the rest of the East Asia region, and you can see it has a total of 34 non-webcomic manga works in total. The larger the sample size in each of these subsets the higher the probability that the effects of publication date, genre, author or work specific traits get evened out. With this in mind these numbers seem to indicate that we could try and compare Southeast Asian manga with Japanese manga as well as manga from other regions. Note, that three regions with too few manga in the corpus have been left out of this table for the sake of brevity. The high number of webcomic format manga in the Southeast Asian part of the corpus also allows us to compare webcomic manga with print format manga. This distinction is well-motivated by preceding research showing that the online publication format has a strong tendency of altering the paneling and layout of comics pages, which is precisely what we will be looking at in this example.

[13]

Among the many annotations available in the TINTIN corpus we have information for the attentional framing types of each panel, a typology introduced by Neil Cohn already in his early work on visual languages. This framework distinguishes between six types of panels based on their contents: macros, containing multiple active entities, monos with a single active entity, micros, which feature only smaller parts of an active entity, amorphic panels containing no active entities, affixing panels, which contain only morphological suffixes, such as the speed lines in the example here, and null panels with no information, usually being a single black or white panel. The “fingerprints” of different stylistic families of comics based on their panel transition distributions, as introduced by McCloud in *Understanding Comics*, is a well-known example of how the different nature of manga paneling and story development can be grasped through such a quantitative summary of a particular aspect of the structure of comics.

[14]

Correspondingly, and not unrelated to the distribution of panel transitions, the distribution of panels by attentional framing types also captures many of the characteristic elements of the structure of paneling and story development that contribute to setting apart different comics stylistic families. Here we have the different regional variations for this distribution for manga, with webcomics separated out. Since only Southeast Asia has a sizable enough sample of webcomics manga, all webcomics manga from other regions – please recall that there are no Japanese webcomics in the corpus – are grouped into the “Other manga (webcomic)” category. Also, as in the previous table, Japan is its own category, separated out from the rest of East Asian manga. Note, that due to the difference in sample sizes and the lengths of works analyzed, we are comparing proportions here not absolute numbers.

[15]

It is not hard to imagine the rough overarching pattern for manga for panels’ attentional framing type distribution based on this diagram, which would prove distinct from other comics styles, if we were to focus on that aspect of the problem.

[16]

However, instead we want to draw attention here to the variation that nevertheless exists between the distributions plotted here. To get a better sense of these differences, let us examine two sets of comparisons.

[17]

First, we have manga from Japan, Southeast Asia and East Asia (again without Japan).

[18]

Interestingly, when comparing all the analysed distributions pairwise with each other, all three of these distributions are statistically different from each other, despite the first two looking very similar to the naked eye. Nevertheless, out of the ten distributions the Southeast Asian one is the closest in similarity to that of the Japanese based on the statistical test results as well.

[19]

As expected, the distribution of panels' attentional framing types for webcomic manga from Southeast Asia has a clearly different distribution from that of the non-webcomic manga from the same region. On the other hand, it is visibly similar to the panel distribution of other webcomics as well as African (non-webcomic) manga.

[20]

This similarity is in fact so strong that based on the pairwise statistical tests these distributions could have come from the same underlying distribution.

[21]

To summarize in one diagram the statistical similarity of all the distributions that proved to be similar to each other as a result of the pairwise comparisons, we have created a graph of these distributions, with the node size representing the sample size for the given region/format, the node color indicating the format (turquoise being webcomic), and the color and thickness of the edges showing the level of similarity between the distributions characterising each of these samples. Note, that the magenta colored edges would indicate a similarity only at – or very close to – the stricter 99% significance level. One important point that leaps at us immediately from this graph is that pairwise statistical similarity between distributions is not a transitive relationship, meaning that despite two distributions both being statistically similar to a third one, they might not be statistically similar to each other. The most striking example of this is the lack of statistical similarity between the distributions characterising South American and African manga, both of which are themselves highly similar to North American manga.

[22]

The strongest clique of nodes forming a fully connected subgraph with all connections among them indicating medium to high levels of pairwise similarity is the one highlighted here made up of other and Southeast Asian webcomic manga, and African and North American manga.

These are, of course, only the first steps in the exploration enabled by the TINTIN corpus regarding manga from Southeast Asia. It is also important to keep in mind that the samples for the respective regions/formats are not representative, and thus a deeper examination of each sample according to genre, publication date, and so on would be the logical next step in such an exploration. Extending the corpus with further representative

works with the help of the MAST annotation tool would also be a very productive avenue to ensure that the results are even more robust. With that being said, we hope to have demonstrated how even the already available data in the TINTIN corpus can be used to generate interesting insights and hypotheses for further exploration and research.

[23]

Next up, let us take a look at the opportunities opened up for studying popular visual media from East Asia, more specifically from China and South Korea with the help of the JVMG knowledge graph,

[24]

which despite its name actually contains a wealth of information on animation, comics and visual novels from these two countries as well, in part reflecting the interest of the communities compiling the source data that are then incorporated.

Here we can see the number of works for which information has been recorded in AnimeClick, which has the great advantage of also containing both genre and content related tags for most of these works enabling a wide range of analyses. There are a larger number of Chinese web animation, and Manhwa (as well as specifically webtoons) in the data.

[25]

What is more, there are also a significant number of creators – both companies and individuals – recorded for these works along with their respective roles in the production process.

Although, once again we need to emphasize that this data is not representative of all Chinese and South Korean animation and comics works, they might be representative of the ones that are internationally the most well-known (since these records originate from an Italian language database). Thus, while caution is always advised for the proper framing of what types of statements can be made based on this data, it does offer a level of depth that warrants potential interest from researchers working on these topics.

[26]

This caveat is less relevant in the case of the data on visual novels from Chinese and Korean language producers found in the JVMG knowledge graph. This data originates from The Visual Novel Database, which is not focused on Japanese games only. In fact, as the late founding creator of the site, Yoran Heling, who left us tragically early this year, emphasized in his presentation at the JVMG Symposium in late February, over the past three years the number of new non-Japanese language releases being added to the database have started to eclipse that of new Japanese language ones. As such the

chances of a high – and potentially even representative – coverage of Chinese and South Korean visual novel releases is definitely a possibility, but would, of course, need further corroboration for more in-depth research work.

Turning to the table itself, we can see here that there are a large number of both Chinese and Korean language producers as well as releases developed and/or published by them in the data.

[27]

To once again highlight an interesting thread that emerges already from this summary overview level of the data, when looking at the ratios between developed and published releases we can see a couple of differences between Chinese and Korean language producers. Chinese language companies are most likely more strongly involved in localization than original production as the number of published releases outnumbers developed ones three to one, and this surplus of published works is chiefly due to translation and localization work. This ratio is much more even for Korean language companies, indicating that on the whole original production is most likely twice as important in their portfolios as localization work. We can also see that this trend is somewhat reversed in the case of individuals, with Chinese language individual producers being more active in the creation of original works than their Korean language counterparts. The overall release numbers, nevertheless, still show a stronger tendency towards original production for Korean language producers and translation-localization for Chinese language producers of visual novel games.

[28]

Thank you very much for your attention, and we hope that these two examples have piqued your interest in the data available in the TINTIN corpus and the JVMG knowledge graph. Please take a look at the available materials on our websites, or contact us directly, if you are interested in working with the data or tools introduced today.