Disseminate your Research: Style Does Matter

Pierre Senellart
The content of your research is the most important thing...
Substance vs Style

- The **content** of your research is the most important thing...  
- ... but the way you **present** it can make the difference.
The content of your research is the most important thing... but the way you present it can make the difference.

Accept or Reject?
Should I read this paper?
Is this worth funding?
The role of writing, presentation, and dissemination

- Serving the substance of your research
- First and semi-conscious lasting impression on an article, a presentation, a work
- Facilitating reading, making your proofs clear, highlighting your results, putting into context, providing supporting material
- In science, style should never go before substance
- All of this is only my personal view 😊
English grammar. Very important, but I am neither a native speaker nor an English teacher.

Writing style. Important as well, but I am not a literature major.

Oral presentation skills. Practice this!

Content, domain-specific. Ask your advisor!
What I will not talk about

**English grammar.** Very important, but I am neither a native speaker nor an English teacher.

**Writing style.** Important as well, but I am not a literature major.

**Oral presentation skills.** Practice this!

**Content, domain-specific.** Ask your advisor!

Examples come from PIKM 2012 articles (except for charts, from CIKM 2013 articles).
Outline

Style Does Matter

Typography of Research Articles
  Don’ts
  Do’s

Visual Display of Quantitative Information

Online Presence

To Go Further
BAD OD TYPGRAPHY IS EVERYWHERE

GOOD TYPGRAPHY IS INVISIBLE / BAD TYPGRAPHY IS EVERYWHERE

Craig Ward
Why typography conventions matters

Typography is the art and technique of arranging type in order to make language visible. Wikipedia
Why typography conventions matter

Typography is the art and technique of arranging type in order to make language visible.  

- Typography has developed since the invention of the printing press.
- A number of typographical conventions have emerged, beyond grammar and language usage conventions.
- Sometimes chosen for readability, sometimes fully arbitrary.
- Because we are used to read printed documents (or documents typeset on a computer screen), we have assimilated these conventions.
- Any deviation from these conventions is noticed and distracts from the actual content.
Should we always respect conventions?

Not necessarily! Some conventions may actually hurt readability of some kinds of typeset material.

Example (U.S. usage)

Periods and commas always go inside quotation marks, even inside single quotes.

The keyword query used in the experiment was "olympics 2012."

or

The keyword query used in the experiment was "olympics 2012".

But:

Know the conventions

Have a good reason to deviate when you do

Be consistent
Should we always respect conventions?

- Not necessarily!

Example (U.S. usage)

Periods and commas always go inside quotation marks, even inside single quotes.

The keyword query used in the experiment was “olympics 2012.”

or

The keyword query used in the experiment was “olympics 2012”.

But:

- Know the conventions
- Have a good reason to deviate when you do
- Be consistent
Should we always respect conventions?

- Not necessarily!
- Some conventions may actually hurt readability of some kinds of typeset material.

Example (U.S. usage)

Periods and commas always go inside quotation marks, even inside single quotes.

The keyword query used in the experiment was "olympics 2012."
or

The keyword query used in the experiment was "olympics 2012".

But:

Know the conventions
Have a good reason to deviate when you do
Be consistent
Should we always respect conventions?

- Not necessarily!
- Some conventions may actually hurt readability of some kinds of typeset material.

Example (U.S. usage)

*Periods and commas always go inside quotation marks, even inside single quotes.*
Should we always respect conventions?

- Not necessarily!
- Some conventions may actually hurt readability of some kinds of typeset material.

**Example (U.S. usage)**

*Periods and commas always go inside quotation marks, even inside single quotes.*

The keyword query used in the experiment was “olympics 2012.”

*or*

The keyword query used in the experiment was “olympics 2012”.
Should we always respect conventions?

- Not necessarily!
- Some conventions may actually hurt readability of some kinds of typeset material.

Example (U.S. usage)

*Periods and commas always go inside quotation marks, even inside single quotes.*

The keyword query used in the experiment was “olympics 2012.”

*or*

The keyword query used in the experiment was “olympics 2012”.

- But:
  - Know the conventions
  - Have a good reason to deviate when you do
  - Be consistent
Be consistent!
Outline

Style Does Matter

Typography of Research Articles
  Don’ts
  Do’s

Visual Display of Quantitative Information

Online Presence

To Go Further
What's wrong with this?

“dangerous” or “non-dangerous”
What’s wrong with this?

"dangerous" or "non-dangerous"

Abuse of typography conventions

- Weird usage of quotation marks: “…”
Several problems remain to be addressed in future research. Additional work is required to improve the integration and exploration of source code repositories and a database’s schema (e.g. parallel processing techniques).
Several problems remain to be addressed in future research. Additional work is required to improve the integration and exploration of source code repositories and a database’s schema (e.g. parallel processing techniques).

Abuse of typography conventions

- Missing comma after “e.g.”
- Improper double spacing after “e.g.”
For both input topics ‘CAD’ and ‘Kashmir’, we observe a very small percentage of users (CAD: 17%, Kashmir: 0.8%) which have the input topic in their bio. But, the
Abuse of typography conventions

- Inconsistent spacing before “%” (the preferred writing is without a space)
- Extraneous space before colons and comma
type conversion, without the need to change the original data. For example it is possible to define Attribute 1 as a sum of Data Field 1 and Data Field 2. It is also possible to use a data field in more than one mapping. Although they can be bound to the same data,
type conversion, without the need to change the original data. For example it is possible to define Attribute 1 as a sum of Data Field 1 and Data Field 2. It is also possible to use a data field in more than one mapping. Although they can be bound to the same data,

Bad line break
- Breaking a line between “Data Field” and “1” hurts readability
What’s wrong with this?

for each $j = 1, 2, ..., \text{the probability } \theta_{ji} \text{ related to the activity } j \text{ are learned with a HMM of states } x \text{ and observations } y,$
for each $j = 1, 2, \ldots$, the probability $\theta_{ji}$ related to the activity $j$ are learned with a HMM of states $x$ and observations $y$.

**Bad line break**

- Breaking a line between “activity” and $j$ hurts readability
What’s wrong with this?

\[ \text{ir} f(k) = \log \left( \frac{n}{\hat{n}} + 1 \right) \ast \log \left( \frac{|E|}{\hat{e}} + 1 \right) \]
What’s wrong with this?

\[ irf(k) = \log \left( \frac{n}{\hat{n}} + 1 \right) \times \log \left( \frac{|E|}{\hat{e}} + 1 \right) \]

- * instead of \(\times\) (or \(\cdot\)) for multiplication
- Improper kerning in the function name, should be: \(irf\)
What's wrong with this?

\[ \text{Dist}_{\text{max}}(s, l, t) = \max \{ \text{Dist}(d_i, s, l, t) | d_i \in D_{s, l}^t \} \]
What’s wrong with this?

\[ \text{Dist}_{\text{max}}(s, l, t) = \max \{ \text{Dist}(d_i, s, l, t) \mid d_i \in D^t_{s, l} \} \]

- *max* should be *max*: this is an operator name, not a variable name.
- Improper spacing around the \(|\) operator; it should be typeset using \(\mid\), not \(|\).
Figure 8: The figure shows the scatter plot for betweenness centrality versus degree centrality for the topic-centric community ‘Kashmir’ (k = 15).
What’s wrong with this?

Figure 8: The figure shows the scatter plot for between-ness centrality versus degree centrality for the topic-centric community ‘Kashmir’ (k = 15).

- The authors have tried tweaking the size of the caption, probably using a \small command, probably to gain space.
- Bad idea to change the default style imposed in the document class, for consistency.
- More importantly, since only part of the paragraph has changed font space, line spacing has not been changed! It is now too large, and no space is saved at all.
What’s wrong with this?

1: for \( t_i \) in \( T \) do
2: \( C_{\text{current}}^{t_i} = \text{search\_bio}(t_i) \)
3: \( U^{t_i}, B^{t_i}, G_{\text{current}}^{t_i} \leftarrow \text{expand}(C_{\text{current}}^{t_i}) \)
4: while Continue\_Condition(\( R_{max}[\ldots] \)) is \text{TRUE} do
What's wrong with this?

1: for $t_i$ in $T$ do
2: $C^t_i_{\text{current}} = \text{search\_bio}(t_i)$
3: $\mathcal{U}^t_i, \mathcal{B}^t_i, G^t_i_{\text{current}} \leftarrow \text{expand}(C^t_i_{\text{current}})$
4: while Continue\_Condition($\mathcal{R}_\text{max}[\ldots]$) is TRUE do

- Bad kerning inside subscripts. One has the impression that $c, u, r, r, e, n, t$ are individual variables that are multiplied.
- “current”, “max” are not variables, they are labels, they should typeset in upright font, not italics.
- Inconsistency in the use of $\leftarrow$ and $=$ for assignment.
What’s wrong with this?

What’s wrong with this?


Weird formatting of bibliography

- Probable improper use of BibTeX syntax ("and" keyword)
- A URL should not be hyphenated when spanning line breaks
What’s wrong with this?

Figure 1: Matches between a program $P$ and a schema $S$. 

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Matches between a program $P$ and a schema $S$.}
\end{figure}
What’s wrong with this?

Figure 1: Matches between a program $P$ and a schema $S$.

Clashing image

- Low-resolution bitmap image
- Inconsistency between image typeface and text typeface
What’s wrong with this?
What’s wrong with this?

Clashing image

- JPEG compression artifacts
Outline

Style Does Matter

Typography of Research Articles
  Don’ts
  Do’s

Visual Display of Quantitative Information

Online Presence

To Go Further
# Punctuation and spacing

<table>
<thead>
<tr>
<th>French</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>. followed by a space</td>
<td>followed by a double space (usually, when full stop)</td>
</tr>
<tr>
<td>, followed by a space</td>
<td>followed by a space</td>
</tr>
<tr>
<td>; followed by a space and preceded with a thin unbreakable space</td>
<td>followed by a space</td>
</tr>
<tr>
<td>: followed by a space and preceded with a thin unbreakable space</td>
<td>followed by a space</td>
</tr>
<tr>
<td>? followed by a space and preceded with a thin unbreakable space</td>
<td>followed by a space</td>
</tr>
<tr>
<td>! followed by a space and preceded with a thin unbreakable space</td>
<td>followed by a space</td>
</tr>
</tbody>
</table>
Dashes and quotation marks

<table>
<thead>
<tr>
<th>\LaTeX{}</th>
<th>French</th>
<th>American English</th>
</tr>
</thead>
<tbody>
<tr>
<td>_ _</td>
<td>petite-fille, 1979-1981</td>
<td>grand-daughter</td>
</tr>
<tr>
<td>_ _ _ _</td>
<td>— Bonjour — Ça va ?</td>
<td>I believed him—since he was convincing—and I followed him.</td>
</tr>
</tbody>
</table>

Il a dit : « Bonjour ».

In British typography, usage of ‘ ’ and “ ” is reversed.

He said, “Hi”, I reckon. “He said, ‘Hi’, didn’t he?”
Pay special attention to line breaks, avoid breaking up **semantic phrases** (a number and its unit, a variable and its type, etc.), so as not to disrupt the flow of reading.

Use nonbreakable spaces (~) where needed:

- Given a set~$X$
- As shown in Sections~\ref{a}, \ref{b}, and~\ref{c}
- The program ran in 23~ms
- As mentioned in~\cite{foobar}

Pay special attention to page/column breaks, **avoid a break after a section heading, avoid widows or orphans**... \LaTeX{} usually good at this for standard classes, but may not be the case with publisher-provided classes.

Use explicit breaks (\pagebreak) as needed, as a final step.
Stick to a stylesheet

- If using Word or a similar WYSIWYG tool, rely on styles, not on explicit formatting instructions (but you should use LaTeX instead 😊)
- In LaTeX, avoid explicit formatting commands (\textbf, \small, \vspace, \, \noindent...); they are occasionally useful but tend to be abused
- Stick to semantic commands and environments provided by the document class (theorems, sections, paragraph headings, etc.)
- If using a publisher-provided document class, read the documentation and carefully follow the instructions
- If no class imposed, rely on modern document classes (scrartcl, memoir, scrbook...) rather than on legacy LaTeX classes (article, report)
Write semantic math

- Define \texttt{macros} for every multiple-letter variable (\texttt{\textbackslash mathit}), and every label (\texttt{\textbackslash mathrm}), operator (\texttt{\textbackslash DeclareMathOperator}) and special notation; this will ensure proper formatting and consistent usage.

- Use \texttt{\textbackslash text} for incorporating \texttt{text fragments} inside equations.

- Use \texttt{amsmath} extensions to break equations on multiple lines, to format matrices, modulos, etc. Read the documentation of this package!

- The way the equation is written should (as much as possible) reflect its \texttt{semantics}, not the way it is displayed.

- Use \texttt{inline} ($ $ or \texttt{\( \textbackslash ( \textbackslash \))}) and \texttt{displayed} (\texttt{\[ \textbackslash [ \textbackslash ]}) modes as required.

- Do \textbf{not} use legacy $$ $$ and \texttt{eqnarray}, poor formatting.
Harmony between figures and text

- Figures should **not stand out**, should not appear as foreign objects inside the document
- Use **vector graphics** rather than bitmaps
- If you have to use bitmaps, go for **high-resolution, lossless compression** (PNG rather than JPG)
- Use the **same font** as in the text, **same font size, same notation**, same line width as in tables, etc.
- Consider producing **\LaTeX code for figures**, using tools such as TikZ (read the documentation!), metapost, asymptote, Inkscape PDF+\LaTeX export, dot2tex, etc. This ensures common look, allows the use of user-defined macros in drawing, etc.
Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information
  Don’ts
  Do’s

Online Presence

To Go Further
We shouldn’t abbreviate the truth but rather get a new method of presentation.

Edward Tufte
We shouldn't abbreviate the truth but rather get a new method of presentation.

Edward Tufte
Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Don’ts

Do’s

Online Presence

To Go Further
### What’s wrong with this?

**Badly formatted numerical data**
- Numbers not aligned on their least-significant digit
- Inconsistent (and uselessly high) number precision

<table>
<thead>
<tr>
<th>action</th>
<th>LO vs LE $\chi^2$</th>
<th>p-value</th>
<th>LO vs CA $\chi^2$</th>
<th>p-value</th>
<th>LE vs CA $\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE</td>
<td>9</td>
<td>0.011</td>
<td>9</td>
<td>0.029</td>
<td>18</td>
<td>0.006</td>
</tr>
<tr>
<td>FV</td>
<td>9</td>
<td>0.011</td>
<td>9</td>
<td>0.011</td>
<td>18</td>
<td>0.001</td>
</tr>
<tr>
<td>LE</td>
<td>13</td>
<td>0.043</td>
<td>6</td>
<td>0.301</td>
<td>27</td>
<td>0.079</td>
</tr>
<tr>
<td>QL</td>
<td>36.563</td>
<td>0.064</td>
<td>45</td>
<td>0.039</td>
<td>45</td>
<td>0.039</td>
</tr>
</tbody>
</table>
What’s wrong with this?

<table>
<thead>
<tr>
<th>action</th>
<th>LO vs LE</th>
<th>LO vs CA</th>
<th>LE vs CA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>$p$-value</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>BE</td>
<td>9</td>
<td>0.011</td>
<td>9</td>
</tr>
<tr>
<td>FV</td>
<td>9</td>
<td>0.011</td>
<td>9</td>
</tr>
<tr>
<td>LE</td>
<td>13</td>
<td>0.043</td>
<td>6</td>
</tr>
<tr>
<td>QL</td>
<td>36.563</td>
<td>0.064</td>
<td>45</td>
</tr>
</tbody>
</table>

**Badly formatted numerical data**

- Numbers not aligned on their least-significant digit
- Inconsistent (and uselessly high) number precision
What's wrong with this?

<table>
<thead>
<tr>
<th>element</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headline</td>
<td>UES</td>
</tr>
<tr>
<td>passage within an article</td>
<td>AA</td>
</tr>
<tr>
<td>List</td>
<td>LI</td>
</tr>
<tr>
<td>Introduction</td>
<td>EL</td>
</tr>
<tr>
<td>Picture</td>
<td>BI</td>
</tr>
<tr>
<td>Info Box</td>
<td>IB</td>
</tr>
<tr>
<td>Charts, tables etc.</td>
<td>IG</td>
</tr>
<tr>
<td>Navigation within Wikipedia</td>
<td>IW</td>
</tr>
<tr>
<td>Other navigation</td>
<td>BA</td>
</tr>
</tbody>
</table>
### What’s wrong with this?

<table>
<thead>
<tr>
<th>Element</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headline</td>
<td>UES</td>
</tr>
<tr>
<td>Passage within an article</td>
<td>AA</td>
</tr>
<tr>
<td>List</td>
<td>LI</td>
</tr>
<tr>
<td>Introduction</td>
<td>EL</td>
</tr>
<tr>
<td>Picture</td>
<td>BI</td>
</tr>
<tr>
<td>Info Box</td>
<td>IB</td>
</tr>
<tr>
<td>Charts, tables etc.</td>
<td>IG</td>
</tr>
<tr>
<td>Navigation within Wikipedia</td>
<td>IW</td>
</tr>
<tr>
<td>Other navigation</td>
<td>BA</td>
</tr>
</tbody>
</table>

#### Badly formatted table

- Inconsistent use of lower and upper case
- Not enough contrast between header and table content
- Useless space in between columns, hurting readability
- Missing comma before “etc.”
What’s wrong with this?

<table>
<thead>
<tr>
<th>Musicians</th>
<th>Films</th>
<th>Actors</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Johnny Cash</em></td>
<td><em>Braveheart</em></td>
<td><em>Al Pacino</em></td>
</tr>
<tr>
<td>Bob Dylan</td>
<td>We Were Soldiers</td>
<td>Robert De Niro</td>
</tr>
<tr>
<td>Elvis Presley</td>
<td>The Passion of the Christ</td>
<td>Robert Duvall</td>
</tr>
<tr>
<td>June Carter Cash</td>
<td>Patton</td>
<td>Marlon Brando</td>
</tr>
<tr>
<td>Willie Nelson</td>
<td>Saving Private Ryan</td>
<td>Kevin Spacey</td>
</tr>
<tr>
<td>Waylon Jennings</td>
<td>Flags of Our Fathers</td>
<td>Ellen Barkin</td>
</tr>
</tbody>
</table>
What’s wrong with this?

Hard-to-read table

- Checkerboard effect: vertical rules and repeated horizontal rules hamper readability (the brain sees the lines more than the content)
What’s wrong with this?

Figure 3: Precision @ $\bar{K}$ Results
What’s wrong with this?

Misleading chart

- Misleading $x$-axis: trying to follow the evolution of the line chart is meaningless
- Mismatch between chart style and rest of the paper

Figure 3: Precision @ $K$ Results
What’s wrong with this?

Hard-to-read chart

3D plots are notoriously hard to read

How to compare bars at different depth?

Pattern fills are distracting to the eye
What's wrong with this?

Hard-to-read chart
- 3D plots are notoriously hard to read
- How to compare bars at different depth?
- Pattern fills are distracting to the eye
Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information
  Don’ts
  Do’s

Online Presence

To Go Further
Beautiful tables: booktabs

- Forget vertical rules: hurt readability
- Forget \hline: inflexible spacing
- Use the booktabs package for tables

\begin{tabular}{cc}
\toprule
T & Y \\
\midrule
a & b \\
b & c \\
b & a \\
\bottomrule
\end{tabular}
In a toto.dat file:

```plaintext
# Convergence results
# fictional source, generated 2008
level  dof  error1  error2  info  grad(log(dof),log(error2))  quot(error1)
1  4 2.50000000e-01 7.57858283e-01 48 0 0
2 16 6.25000000e-02 5.00000000e-01 25 -3.00000000e-01 4
3 64 1.56250000e-02 2.87174589e-01 41 -3.99999999e-01 4
4 256 3.90625000e-03 1.43587294e-01 8 -5.00000003e-01 4
5 1024 9.76562500e-04 4.1941738e-02 22 -8.49999999e-01 4
6 4096 2.44140625e-04 1.69802322e-02 46 -6.90000001e-01 4
7 16384 6.10351562e-05 8.20091159e-03 40 -5.24999999e-01 4
8 65536 1.52587891e-05 3.90625000e-03 48 -5.35000000e-01 3.99999999e+00
9 262144 3.81469727e-06 1.95312500e-03 33 -5.00000000e-01 4.00000001e+00
10 1048576 9.53674316e-07 9.76562500e-04 2 -5.00000000e-01 4.00000001e+00
```
## Beautiful tables: pgfplotstable (2/2)

<table>
<thead>
<tr>
<th>level</th>
<th>Dof</th>
<th>$e_1$</th>
<th>$e_2$</th>
<th>info</th>
<th>$\nabla e_2$</th>
<th>$\frac{e_1^{(n)}}{e_1^{(n-1)}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>$2.5^{-1}$</td>
<td>$7.58 \cdot 10^{-1}$</td>
<td>+48.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>$6.3^{-2}$</td>
<td>$5.00 \cdot 10^{-1}$</td>
<td>+25.0</td>
<td>–0.3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>$1.6^{-2}$</td>
<td>$2.87 \cdot 10^{-1}$</td>
<td>+41.0</td>
<td>–0.4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>256</td>
<td>$3.9^{-3}$</td>
<td>$1.44 \cdot 10^{-1}$</td>
<td>+8.0</td>
<td>–0.5</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>1 024</td>
<td>$9.8^{-4}$</td>
<td>$4.42 \cdot 10^{-2}$</td>
<td>+22.0</td>
<td>–0.85</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>4 096</td>
<td>$2.4^{-4}$</td>
<td>$1.70 \cdot 10^{-2}$</td>
<td>+46.0</td>
<td>–0.69</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>16 384</td>
<td>$6.1^{-5}$</td>
<td>$8.20 \cdot 10^{-3}$</td>
<td>+40.0</td>
<td>–0.52</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>65 536</td>
<td>$1.5^{-5}$</td>
<td>$3.91 \cdot 10^{-3}$</td>
<td>+48.0</td>
<td>–0.54</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>262 144</td>
<td>$3.8^{-6}$</td>
<td>$1.95 \cdot 10^{-3}$</td>
<td>+33.0</td>
<td>–0.5</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>1 048 576</td>
<td>$9.5^{-7}$</td>
<td>$9.77 \cdot 10^{-4}$</td>
<td>+2.0</td>
<td>–0.5</td>
<td>4</td>
</tr>
</tbody>
</table>
Choose the right chart for the job

- Take time to decide **which chart is appropriate:**
  - Which data points? What is the message?
  - Linear, categorical, or logarithmic axes?
  - Error bars?

- Present experimental results **honestly**, but in a way that highlight your contributions

- **Avoid** any form of clutter or visual gimmick, make it as readable as possible

- For articles, ensure readability when printed in black&white

- Consider using less common visualizations **when appropriate** (radar chart, box plot...). Be both creative and readable.
\begin{tikzpicture}
\begin{loglogaxis}[xlabel=Cost,ylabel=Error]
\addplot[color=red,mark=x]
coordinates {
(5,8.31160034e-02)
(49,7.40715288e-03)
(129, 2.10192154e-03)
(769, 1.62269942e-04)
(1793, 4.44248889e-05)
(4097, 1.20714122e-05)
(9217, 3.26101452e-06)};
\legend{Case 1}
\end{loglogaxis}
\end{tikzpicture}
Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Online Presence
  Do’s

To Go Further
If it doesn’t exist on the Internet, it doesn’t exist.

Kenneth Goldsmith
Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Online Presence

Do’s

To Go Further
Put preprints online

- As soon as a paper is done, put it online, in one of these forms:
  - On your Web page
  - As a technical report of your institution
  - As a preprint on arXiv/CoRR

- State very clearly in the paper and in references to it that it’s an unpublished preprint

- Advantages:
  - People will know more easily about your work, will cite it, will reuse it
  - This can avoid getting scooped
  - Both science and your career will advance faster!

- Not in the following cases:
  - Double-blind review (though reviewers do not usually care)
  - Patent pending
Maintain published papers online

- Once your paper is accepted for publication, make sure it remains online
- Your Web site (if it is stable), arXiv, publication repositories, are suitable stores
- Same advantages as before!
- **Avoid** having very different versions online
- Sometimes publishers are not happy with this, but:
  - Some publishers directly allow you to do this (including ACM since recently); prefer them!
  - Many publishers allow an almost-final version to be put online
  - If you are a UK government employee or a US federal employee, you can always do this by law
  - Open access to science deserves a fight!
Put code and datasets online

- Computer science is special in that it should be easy to repeat experiments; you just need to share your code and datasets
- Even if the code is not well-commented, better than nothing! http://matt.might.net/articles/crapl/
- You may want to keep it proprietary: but will it really be more used or more valuable than if you open-source it?
- Will ensure reuse of your research, further developments
- Clear advantage during the reviewing process

Example

RoadRunner (Creszenzi et al., 2001): early work on Web data extraction. Not the first, not the best, many refinements since. Remains the most cited reference in the area (1024 citations!) because the system (and datasets) is freely available online.
Outline

Style Does Matter

Typography of Research Articles

Visual Display of Quantitative Information

Online Presence

To Go Further


Merci.