Agenda

→ Introduction to the problem and project
→ How much is being shared?
→ What is being shared?
→ CourseVillain today
→ Q&A
Crowd-sourced plagiarism

Students are exchanging coursework through crowd-source platforms like:

- Advertises as host of “course specific study materials,”
- boasting “25 million course-specific study materials” from 11,000 institutions.

In practice, host of student-shared coursework and course materials ranging from syllabi to answer sheets = crowd-sourced plagiarism.
Crowd-sourced plagiarism

Students are exchanging coursework through crowd-source platforms like:

Advertises as host of “course specific study materials,”

boasting “25 million course-specific study materials” from 11,000 institutions.

In practice, host of student-shared coursework and course materials ranging from syllabi to answer sheets = crowd-sourced plagiarism.

ERAU 11/1/20: 103,791 docs.
How much is being shared?

Internet facilitates new iterations of producing / sharing information, including coursework.

- **Networked information economy**
- **Affinity spaces**

(Octaue spaces + clear participation + lack of rigor) = No real understanding of the scale or scope

**Attempt to track and measure the volume of uploaded academic material.**
Course Villain

Custom search engine, built by researchers and students.

- Designed to search CourseHero.com’s API
- Search every 12 hours
- Flag 20 ‘most recent’ matches
- Email notifications
- Custom query terms

An automated approach to an overwhelming issue.
CourseHero.com and Google Alerts

Established “Google Alerts” for uploads/changes regarding ERAU content, compared against CourseHero.com search engine: 12/19 – 6/19

<table>
<thead>
<tr>
<th>Course</th>
<th>Difference in artifacts</th>
<th>Google Alert notifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH 111</td>
<td>+ 3,357</td>
<td>5</td>
</tr>
<tr>
<td>ECON 210</td>
<td>+ 2,794</td>
<td>0</td>
</tr>
<tr>
<td>MGMT 221</td>
<td>+ 1,728</td>
<td>1</td>
</tr>
</tbody>
</table>
Course Villain experiment

Scans CourseHero.com twice daily for 20 most recent documents matching query terms.

➔ 55 reports / 836 “matches”

➔ 360 unique documents

➔ 20 matching university content

➔ 17 matching queries

➔ 16 MATH111 + 1 ECON210

* Lots of noise, repeat ‘hits,’ and removed content *
Conclusions

Online coursework exchange is clearly problematic

Very difficult to track

But definitely not impossible

Primarily a technical challenge

... and it *needs* to be addressed

Highlighted new directions

Revealed greater nuances
What is being shared?

“Plagiarism” usually found to be common, relatively frequent practice.

- Self-reported results: ~ 25% students admit
- Similarity detection results: 25-77% assignments exhibit issues

But notable methodological problems persist.

- Self-reporting on misconduct lacks reliability
- Similarity detection is still subjective

Need for an asymmetric approach to plagiarism
Course ‘compromise metric’

“Plagiarism” and “academic dishonesty” are too nuanced and complex to measure effectively.

Observing a cross-section of what academic material is shared online develops a novel understanding of academic misconduct that gauges the vulnerability such behavior.

\[ \text{“medium value artifacts”} \; + \; \text{“high value artifacts”} \]

total number of course artifacts
## ‘Compromise metric’ results

<table>
<thead>
<tr>
<th>Course</th>
<th>Compromise Metric</th>
<th>Low Value</th>
<th>Medium Value</th>
<th>High Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMN 330</td>
<td>100%</td>
<td></td>
<td>Discussion Questions (60%)</td>
<td>Tests, Papers (40%)</td>
</tr>
<tr>
<td>WEAX 201</td>
<td>44%</td>
<td>Notes (56%)</td>
<td>Homework, Discussion Questions (44%)</td>
<td></td>
</tr>
<tr>
<td>ENGL 123</td>
<td>31%</td>
<td>Notes, study documents (69%)</td>
<td>Discussion Questions (30%)</td>
<td>Papers (1%)</td>
</tr>
<tr>
<td>RSCH 202</td>
<td>56%</td>
<td>Notes (38%)</td>
<td>Homework, Discussion Questions (41%)</td>
<td>Quizzes, Papers (15%)</td>
</tr>
<tr>
<td>MATH 111</td>
<td>71%</td>
<td>Misc. Documents (33)</td>
<td>Homework, Discussion Questions (50%)</td>
<td>Quizzes, Papers (17%)</td>
</tr>
<tr>
<td>UNIV 101</td>
<td>0%</td>
<td>Notes (100%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHYS 102</td>
<td>46%</td>
<td>Misc. Documents (54%)</td>
<td>Homework, Discussion Questions (46%)</td>
<td></td>
</tr>
</tbody>
</table>

*Results do not provide a clear or direct measurement of plagiarism or misconduct.*
Conclusions

Academic misconduct may be more prevalent than self-reported, but towards lower end of similarity detection.

Nearly half of the surveyed artifacts posed a clear danger to their courses’ academic integrity.

Problematic coursework exchange is slightly more prevalent among sampled STEM subjects than others.

Students are exchanging a significant degree of problematic, potentially dangerous coursework online.
Course Villain as a research tool

Methods demonstrate a novel approach to monitoring academic misconduct behaviors.

Monitoring the online exchange of coursework is a promising and relevant means of better understanding the practice of students intentionally submitting the work of others as their own – a keystone of plagiarism and other kinds of misconduct.

Replicable, Aggregate, Data-Driven, Current
Course Villain 3.0

½ web application, ½ desktop application

Same custom query, twice daily scanning/reporting

- Filtered for University content
- Match ID#, doc. name, and doc. type saved to database for record and comparison
  - Repeat matches are eliminated
- First search organized by relevancy; then, by recency

View matching documents in application

- Ignore match, remove from database
- Auto-populate “Copyright Infringement” form
CourseVillain’s future

Pursue external funding and partnerships

Academic content management

- Holistic triangulation
- Secure intellectual property
- Proactive instructional design
- Game and deter the platforms
Thank you!

Please contact us with any questions, comments, or collaborative ideas.

Zac Dixon
dixonz@erau.edu

Kelly George
georged8@erau.edu
References


