Experiences from teaching Introduction to Web Science

Rene Pickhardt
Institute for Web Science and Technologies,
University of Koblenz-Landau, Germany
rpickhardt@uni-koblenz.de

Steffen Staab
Institute for Web Science and Technologies,
University of Koblenz-Landau, Germany
staab@uni-koblenz.de

ABSTRACT

Over the past three years we annually tought the master level course *Introduction to Web Science*. The course was designed as a Massive Open Online Course (MOOC) and consists of 24 lectures of fiveunits each. A unit consists of a video with a length of up to 10 minutes, some learning goals and two to five multiple choice quizes as well as a further reading section. In this text we present our current curriculum in Section 1 share some insights about problems and difficulties that we experienced c.f. section 2. Finally we want to present open questions and tasks for the community in section 3.

1. OUR CURRENT CURRICULUM

Over the past years our curriculum changed every time we taught the course[1][2][3]. For the last two years some parts of the course were held as a MOOC.[4] Also in the next term we will change the curriculum again it looks like we are somehow converging towards a stable result. Since our anticipated curriculum by now follows to 70% the structure of the article: *A framework for Web Science* by Tim Berners-Lee et.al. [2] we will state some of the overlaps and differences. From a structural point of view we leave out the first two sections about methodology of Web Science and the introduction resulting in a course that has four parts:

1.1 Technical foundations of the Web

Our main focus is the architecture of the Internet including topics like Ethernet, IPv4, Transmission Control Protocol and the Domain Name System. These topics are presented from a technical and historical perspective so that our students understand the basic dataflow and design principles of the Internet which have an influence on questions concerning web governance. Additionally in this section we introduce the notion of an URI, the Hypertext transfer protocol with its ability to do content nagotiation and XML as a more semantic perspective of HTML. Since we already spend eight lessons on the technical foundations we leave out more topics from the Semantic Web or peer 2 peer systems. With our goal to extend this class to a two semester course these topics would certainly have to be included. The main question we are following in this part of the course could be paraphrased in one word as: **How?**

1.2 Modeling the emerging Web Properties

Spending six lessons in this section we investigate the structure and topology of the Web Graph from a macroscopic point of view. The touched topics can be seen as a subset of a network theory class c.f. [3]. In particular we put an emphasize on the dynamic nature of the Web Graph and its value as a network from an economical point of view. We also present generative models from a microscopic perspective. Most important is the model of preferental attachment [1] since as a random process it can be seen to be the same as urn models which can be used as generative models for words on the web. Students should understand how the structure of the Web influences processes like the flow of information from the example of meme spreading [5]. Fields of interest are: Graph theory, Statistics, Physics, Linear Algebra and Modelling. The main question we are following in this part of the course could be paraphrased as: **What?**

1.3 Behavioural Models

Another six lessons are devoted to this section where we investigate the behavior of users on the web. The main disciplines that we are touching are economics, rational choice theory, sociology and psychology. We investigate herding behavior [4] of Web users as well as filter bubble effects that arise from personalization. We also look at the theories of social capital and take search engines as an example that are able to gain value from various kinds of social capital that is on the web. We look at Googles online advertising system from a game theoretic as well as a multistake holder point of view. Finally we touch the current trends of crowd sourcing and social machines. The main question we are following in this part of the course could be paraphrased as: **Why?**

[1] http://west.uni-koblenz.de/de/studium/lehrveranstaltungen/ws1213/webscience
[2] http://west.uni-koblenz.de/de/studium/lehrveranstaltungen/ws1314/introduction-to-web-science/WebScience
[3] http://west.uni-koblenz.de/de/studium/lehrveranstaltungen/ws1415/introduction-to-web-science
[4] https://en.wikiversity.org/wiki/Web_Science

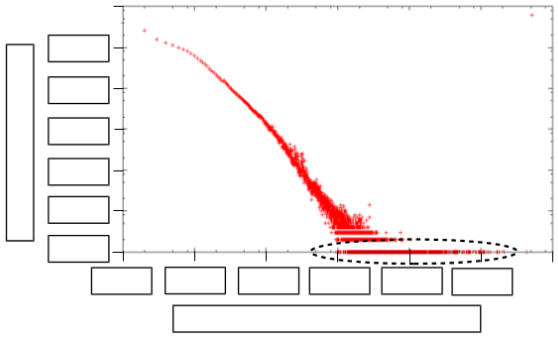
1.4 Web and Society

The final four lessons of our course are spent in this section. Though the overlap with Web governance exists it is not quite the same. Here we discuss topics like copyright from the perspective how a globally connected cyberspace on top of a technical system follows very different paradigms than the regional laws of countries or cultural spaces. From a similar perspective we look at privacy, security, trust and net neutrality. These lessons are mainly to present a broad picture of the Web from a very high level to our students. This section would probably be best described by: **Who?**

2. PRESENT PROBLEMS

Obviously we cannot expect from our students to be aware of all the methods from the fields that we touch in this interdisciplinary curriculum. While being in the classroom we constantly wonder weather we should explain the technical aspects of some methodology or rather touch the topic from a bird eye's perspective and focus on the phenomena? Closely related to this problem we sometimes just have wrong expectations about the prior knowledge of our students. Take the following exam question as an example for these two problems.

The following diagram depicts the indegree distribution of German Wikipedia articles. Please lable the axis correctly. What is the value of the points circled with a dotted line? Roughly sketch the outdegree distribution within the diagram? How are distributions like that called and what kind of a plot do you see?



On average our students received only 43% of the available points. From a total of 23 students not a single student was able to complete all tasks correctly. Though each task was completed correctly more than once (demonstrating that the question was not completely out of scope). Only five students managed to write indegree on the x-axis combined with fequency or probablity on the y-axis. More frequently the answer was given the other way around. Even when putting a logarithmic scale to the axis (which did not happen too often) students frequently stated that the value of the points in the dotted area would be 0 instead of 1. The outdegree distribution frequently showed a growing curve or what we perceived as even worse flipped the indegree distribution so that one could see a big fanout for small x-values and large y-values converging towards a straight declining line for higher x-values.

We perceived this feedback as shocking. Even though we spend quite some lecture time on reading and interpreting diagrams and also had exercises focusing on reading diagrams our students struggled with this question. Our lesson learnt is that while for the research and scientific community it is most interesting to focus on the connections of all the multidisciplinary fields and draw interesting conclusions from them for education we need to focus carefully at one method from one discipline at a time and teach the students technical details rather than looking at the topics from a bird eye's perspective.

Another problem that we experienced is the fact that it is hard to give a clear structure to the topics. Since the field is so complex due to its interdisciplinary nature it allows to study a topic from various perspectives. Even for us as teachers it is not always clear where to place the focus when looking at a certain topic. We assume that for the learner this must be even more confusing. While more and more programs and courses are emerging it seems like they have diferent learning goals and structures. While the already mentioned primer is a starting point it was neither meant to be a strucutre for education nor is it complete.

3. OPEN TASKS FOR THE COMMUNITY

Most importantly it would be great to share our experiences and combine our silo efforts to a unified structure and precise learning goals for a Web Science curriculum. The ultimate outcome would be a book combined with exercises and slides that could be used as a guide for other teachers and would also help students to study the material from one authorative resource. Since our material is under an open license we suggest to collaborate on wikiversity a project of the non profit wikimedia foundation to bring open and free educational resources to the world. Maybe interested educators of different universities could make a seperate three or four day workshop with hands on sessions to combine efforts and create this structure which could eventually lead to a joint book.

4. REFERENCES

[1] Albert-Laszlo Barabasi, Reka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1-4):69 – 77, 2000.

[2] Tim Berners-Lee, Wendy Hall, James A Hendler, Kieron O'Hara, Nigel Shadbolt, and Daniel J Weitzner. A framework for web science. *Foundations and trends in Web Science*, 1(1):1–130, 2006.

[3] Easley David and Kleinberg Jon. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.

[4] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856, 2006.

[5] Lillian Weng, Alessandro Flammini, Alessandro Vespignani, and Fillipo Menczer. Competition among memes in a world with limited attention. *Scientific reports*, 2, 2012.