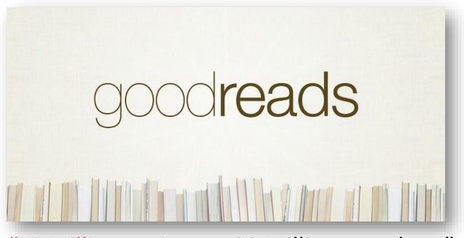
Absorbed in Goodreads

A Computational Approach for the Study of Online Social Reading

Simone Rebora (University of Göttingen) simone.rebora@phil.uni-goettingen.de
Piroska Lendvai (University of Göttingen) piroska.lendvai@uni-goettingen.de
Moniek Kuijpers (University of Basel) moniek.kuijpers@unibas.ch

The Idea

Text mining the reviews published on social reading platforms (such as Goodreads) through the Story World Absorption Scale (SWAS)



"80 Million reviews – 80 Million members"
(www.goodreads.com: *internal statistics*)

The SWAS

Attention (α: .905)

A1: When I finished the story I was surprised to see that time had gone by so fast.

A2: When I was reading the story I was focused on what happened in the story.

A3: I felt absorbed in the story.

A4: The story gripped me in such a way that I could close myself off for things that were happening around me.

A5: I was reading in such a concentrated manner that I had forgotten the world around me.

Transportation (α: .902)

T1: When I was reading the story it sometimes seemed as if I were in the story world too.

T2: When reading the story there were moments in which I felt that the story world overlapped with my own world.

T3: The world of the story sometimes felt closer to me than the world around me.

T4: When I was finished with reading the story it felt like I had taken a trip to the world of the story.

T5: Because all of my attention went into the story, I sometimes felt as if I could not exist separate from the story.

Emotional Engagement (α: .914)

EE1: When I read the story I could imagine what it must be like to be in the shoes of the main character.

EE2: I felt sympathy for the main character.

EE3: I felt connected to the main character in the story.

EE4: I felt how the main character was feeling.

EE5: I felt for what happened in the story.

Mental Imagery (α: .795)

MS1: When I was reading the story I had an image of the main character in mind.

MS2: When I was reading the story I could see the situations happening in the story being played out before my eyes.

MS3: I could imagine what the world in which the story took place looked like.

(Kuijpers et al., 2014)

The Goodreads Reviews

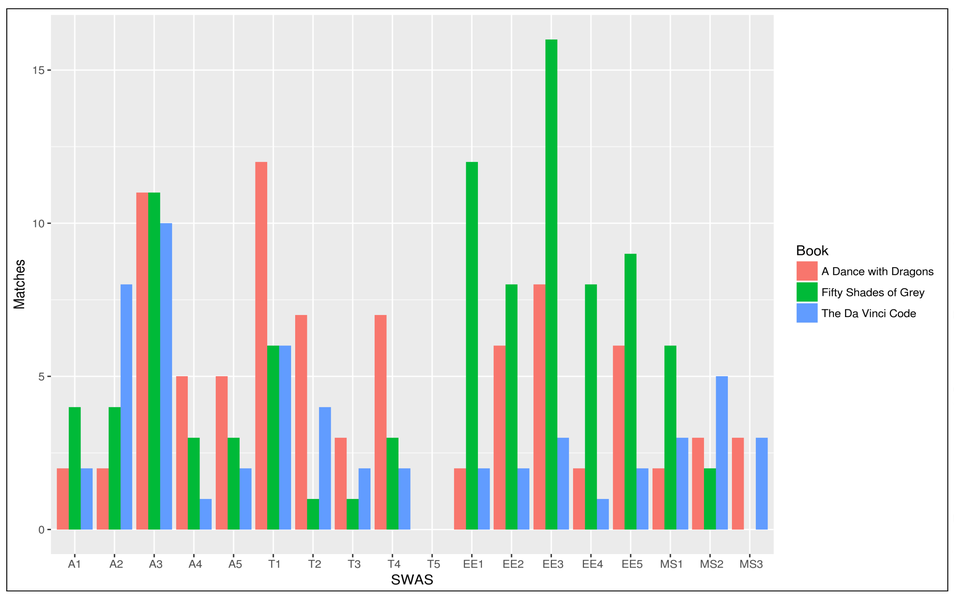
"You know, after reading the book. I almost ripped out my hair from my head. Shouting like a crazy fitting is happening. I'm so absorbed in the world Martin produced out of his wits siguro). The characters were hell of a bomb after bombs. Jon, poor Jon! I nearly cried out to what happened to him in his last POV! It was damn, shit, and nasty things I'd said couldn't claim out what happened! And oh boy! My little Tyrion the giant o' Lan… Who would have thought that she will … new girl! Appropriate of age, the inno… and especially, appropriate of the height herself!"

A3: I felt absorbed in the story

EE3: I felt connected to the main character in the story

"Sometimes I really think I want to punch George R.R. Martin in the face. Really! If I were to ever meet him, it would be a huge emotional struggle to decide whether to cry all over him and leap into his arms thanking him for creating such a marvelous book series...or punching him in the face for making me wait so long in between books. Really. It's a toss up at this point. Because honestly, this book is bloody brilliant! …htaking character …most detailed and …l really felt such a deep connection to certain characters, really invested in their stories and their fates."

Preliminary Manual Analysis



180 reviews of three blockbuster novels.

241 matches between Goodreads reviews and SWAS statements

Results were cross-checked by two annotators

The Corpus

	Fantasy	Romance	Thriller
Reviewed Books	1,243	1,237	1,156
Reviews	735,156	1,037,274	890,420
Sentences (split using Tokenizers)	9,123,531	13,393,344	6,988,237
Tokens	180,858,806	250,262,792	137,549,158

The Goal(s)

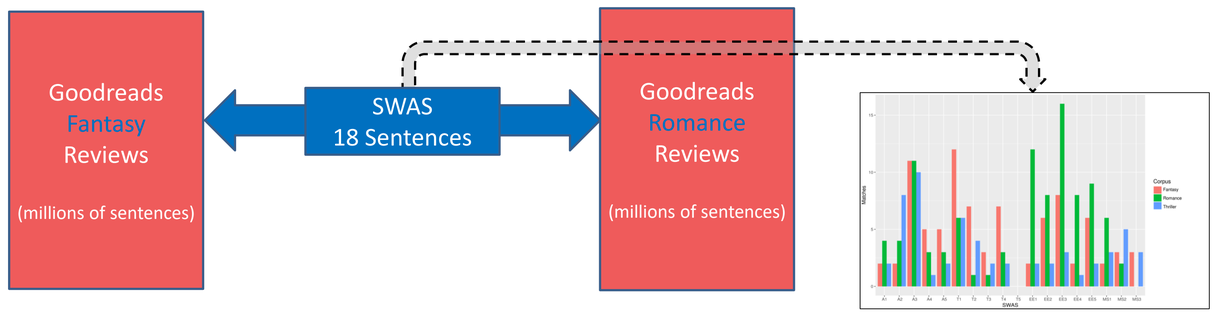
Provide evidence for the theoretical assumptions that…
- fantasy's "worldbuilding" (Wolf, 2012),
- romance's "empathic identification" (Koopman & Hakemulder, 2015),
- or thriller's "embodied simulation" (Gallese, 2007)

…enhance absorption

But also: show differences between reader's gender
…or the effects of different rethorical, conceptual, or contextual features of novels

Validate the SWAS (which SWAS statements match more frequently with Goodreads reviews?)

The Procedure: Automated Pairing of Sentences



Tools Evaluation (Two Approaches)

Text Reuse Detection

TRACER
TEXT REUSE DETECTION MACHINE

Textual Entailment Recognition

EOP
Excitement Open Platform for Recognizing Textual Entailment

	Precision	Recall	F1
TRACER	0.0159	0.2822	0.0301
EOP (untrained)	0.0441	0.0954	0.0604
EOP (trained)	0.0178	0.4917	0.0344

"Ground truth" corpus are the 180 annotated reviews
EOP training was evaluated via 10-fold cross-validation

Producing More Training Material

"Mining Goodreads: a text similarity-based approach to measure reader absorption"
PI: Moniek Kuijpers, Univ. of Basel

- 18-months project, funded by SNSF
- Planned starting date: 1st December 2018
- 4 annotators working on the Goodreads corpus

Pre-Processing the Corpus (a BOW Approach)

- Build a dictionary
  based on the most frequent words
  in the SWAS
  and in the 241 matched sentences*

- Calculate an «absorption score»
  for each review
  and annotate the highest-ranking reviews

Word	Freq
~~book~~	~~34~~
like	22
world	19
~~books~~	~~13~~
characters	12
~~even~~	~~12~~
~~read~~	~~12~~
feel	11
felt	11
~~story~~	~~11~~
completely	10

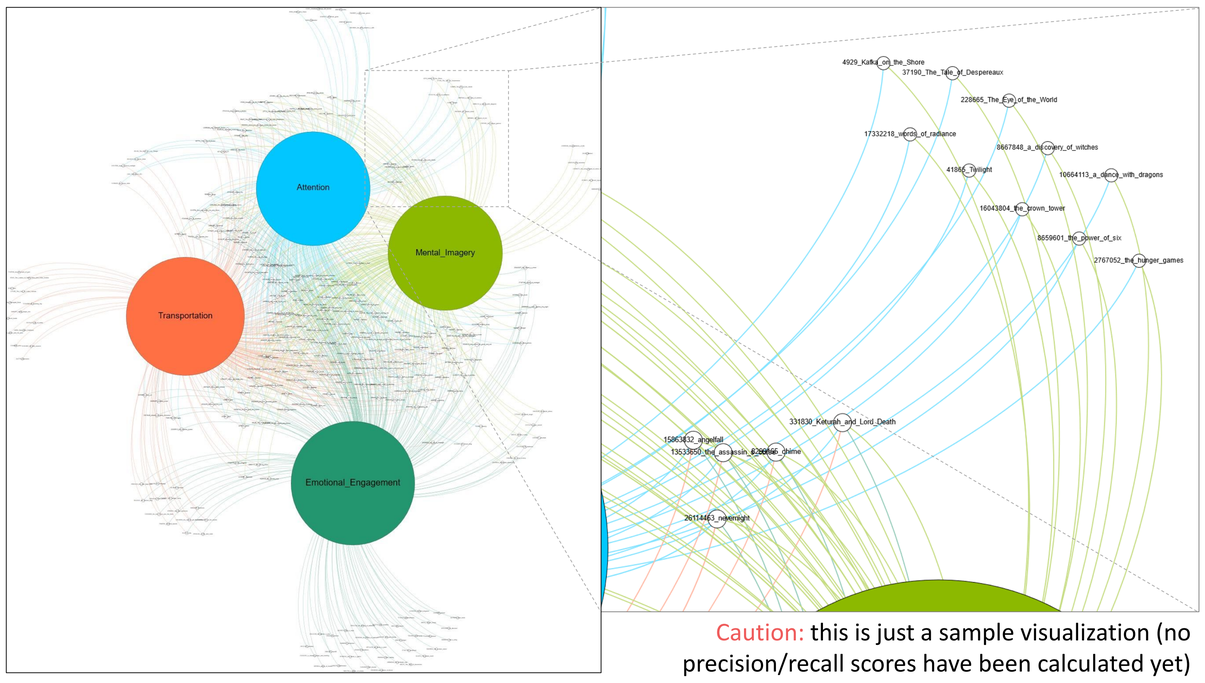
*The list was refined manually

Word	Freq
like	22
world	19
characters	12
feel	11
felt	11
completely	10
love	9
time	9
right	8
character	7
heart	6

- Expand the dictionary through word embeddings
  using word2vec in Gensim
  trained on the Goodreads Corpus (400M tokens)
  on top of the GoogleNews model
  (https://github.com/mmihaltz/word2vec-GoogleNews-vectors)

review_id	Attention	Transportation	Engagement	Imagery
1	0,13	0,13	0,11	0,15
3	0,03	0,02	0,03	0,03
4	0,08	0,09	0,07	0,08
5	0,06	0,06	0,05	0,07
6	0,18	0,18	0,12	0,18
8	0,06	0,07	0,06	0,06
9	0,05	0,05	0,03	0,05
10	0,16	0,16	0,10	0,15

But also… Visualize the Results!



Caution: this is just a sample visualization (no precision/recall scores have been calculated yet)