

# Physics of Cheating

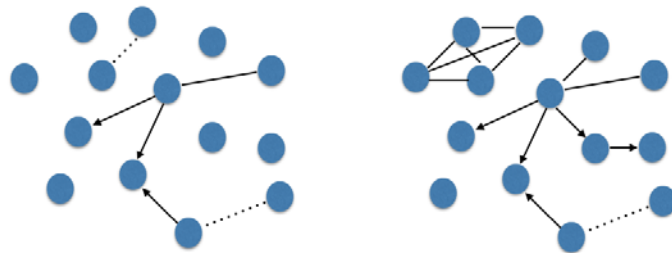
## 1. Information on Research Team

Name	Student ID	Department	Career	Role
이상진	20138073	물리학과	석박통합과정	연구원
한상은	20138116	물리학과	석박통합과정	연구원
이진엽	20165257	수리과학과	박사과정	연구원

## 2. Project Summary

Objective	<p>대학원에서 조교의 업무로서 학생들의 숙제를 채점하다 보면 학생들의 숙제 간에 강한 상호 연관관계가 자주 관측 된다. 우연으로 치부하기엔 이러한 사건들이 빈번하게 발생하지만, 일반적으로 직접적인 증거가 부족하기 때문에 부정행위로 여길 수만은 없는 일이 자주 발생한다. 이 점에 착안하여 숙제간의 상호 연관성에 관한 연구를 진행하고자 한다. 조사과정에서 숙제와 관련된 다양한 변수들과 숙제의 베낌 사이의 상관 관계를 기술할 수 있는 의미있는 측도(measure)를 개발하고, 그 측도를 기반으로 주어진 변수들로 구성할 수 있는 숙제공간을 정의하고 그 공간에서 발생할 수 있는 상전이를 공부해보고자 한다. 이를 통해 네트워크 및 다른 통계 현상 또한 이해할 수 있는지 공부해보려 한다. 그리고 여기서 그치지 않고 이를 이용하여 다른 사회 현상 등에도 적용하여 기존과는 다른 의미있는 해석을 시도해본다.</p>
Description	<p><b>측도개발</b></p> <p>학생들의 숙제를 채점하다 보면 비슷한 유형의 답안지를 자주 보게 된다. 제출한 숙제를 살펴보면, 논리 전개 방식, 단어, 표기법 등 다양한 요소로부터 어떤 학생과 어떤 학생의 숙제가 비슷하거나, 학생들이 숙제를 할 때 함께 토론하였거나 같은 출처의 웹사이트를 사용하는 등의 이유로 제출한 숙제의 연관성이 있을 것이라 생각할 수 있다. 건전한 토론 문화와 적절한 검색 능력은 21세기 교육의 또 다른 면모라고 할 수 있지만 학생들이 자신의 언어로 이해하지 않고 그저 베끼기 식의 숙제를 수행한다면 이는 교과목 학습에 장애를 초래할 것이다. 이 점에 착안하여 숙제의 난이도, 숙제를 하는 학생의 수, 학생들 사이의 거리, 조교의 성향 등 여러 변수를 고려하여 숙제 사이의 상관관계를 수학적으로 엄밀하게 측정할 수 있는 수학적인 측도(measure)를 만들고자 한다.</p> <p><b>통계물리적 접근</b></p> <p>새로이 고안한 측도를 기반으로 주어진 변수를 기저로 하는 일종의 힐베르트 공간(Hilbert space)을 구성한 뒤 우리가 알고 있는</p>

물리적 모델 혹은 수학적 모델로 대응이 가능한 지 조사해 본다면 거시적 혹은 직감적으로는 보이지 않았던 숙제 사이의 관계를 더욱 명확히 볼 수 있을 것이라 예상한다. 이론 물리적 접근과 수학적 엄밀함을 기반으로, 다체문제를 기술하고자 하는 유효 해밀토니안(effective Hamiltonian)을 적는 방식으로 접근해보고자 한다. 이러한 연구를 통해 제안된 모델과 변수들로부터 실제 응집물리 혹은 다체계에서 벌어지는 상전이 현상을 숙제 공간에서도 예상해볼 수 있고 그것이 관찰될 것이라 예상한다. 이러한 모델을 기반으로 학생들의 숙제 시스템을 손으로 풀기는 어려운 통계 문제나 시뮬레이션으로서도 의미가 있는 연구라 생각한다.



Easy problem                      Hard problem  
그림1. 문제의 난이도에 따른 학생들의 연결관계

### 머신러닝과 네트워크를 이용한 접근

이러한 문제에 접근하기 위해 우리는 학생들을 네트워크로 생각할 수 있다. 숙제를 할 때에 이 네트워크에서 숙제에 관한 정보가 상호 영향을 주면서 이동한다고 볼 수 있다. 네트워크에 관한 연구에 따르면, 물리적 라우터와 웹사이트의 논리연결 혹은 인맥관계도에서 상호 연결관계는 멱급수분포(power law distribution)를 따름이 알려져있다. 이와 같이 학생들의 인맥 혹은 숙제 정보 공유 연결망은 멱급수 분포를 따르지 않을까라는 추측을 해볼 수 있다. 즉, 일반적으로 모든 학생이 숙제에 관해 깊이 생각하는 것은 아니기 때문에, 학생들의 숙제 네트워크를 모델링 할 때에 숙제에 관한 정보가 일정 출처- 열심히 하는 학생 혹은 웹에 공유된 자료- 를 가지고 퍼져나가는 모델을 생각해 볼 수 있을 것이다. 이러한 시스템이 유용한 이유는 한 학기 동안 학급내에서 인맥관계는 크게 변화하지 않을 것으로 기대되기 때문에 학생들간의 네트워크는 고정되는 것으로 가정할 수 있고 이를 이용해서 문제를 조금 더 간단하게 만들 수 있다. 이를 기반으로 인맥 변화라는 자유도를 추가해본다면 더 다양한 현상을 기술할 수 있을 것이라 생각한다. 이러한 접근을 위해서는 우리가 처리해야할 데이터의 양이 매우 많아지게 된다. 이러한 문제에 대해 머신러닝을 이용한 전략을 사용할 수 있다. 머신러닝 전략은 큰 데이터 집합 내의 패턴을 파악하는데 유용하다. 이러한 맥락에서 위에서 제시한 현상론적이고 해석적인 접근방법과는 상호보완적으로 머신러닝을 이용해서 학생들의 숙제 데이터를 분석한다면 보다 정확한 함수를 얻는데 큰 도움이 될 것이다. 학생수가 30명만 있다고 해도 총  $30C_2 (=435)$ 개의 짝(pair)이 존재하기 때문에 충분히 커다란 데이터를 갖고 있다고 할

수 있다. 따라서 머신러닝을 활용하여 자료를 분석하는데 어려움이 없다고 생각한다.

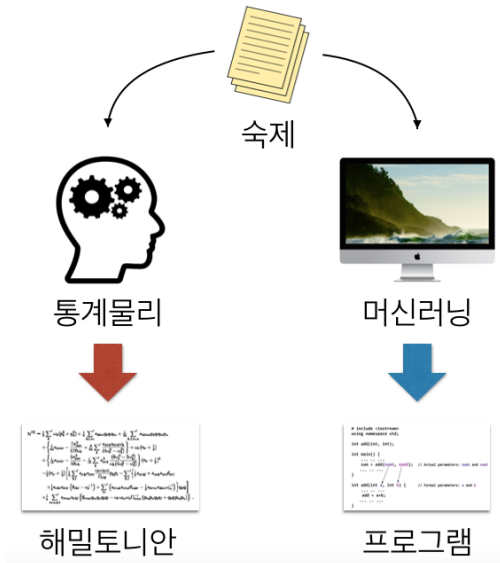


그림2. 연구개요

정리하자면 제안된 문제를 두 가지 방법으로 동시에 해결해보려 한다. 주어진 문제를 알려진 해밀토니안들과 약간의 변형으로 이해해보려는 시도와 모든 가능성을 효율적으로 고려해가는 방법으로 접근해보고자 한다.

Novelty / unusualness

**Unusualness**

숙제간의 상호작용 혹은 숙제 공간의 구조는 많은 사람이 궁금해 왔지만, 문제 자체가 어렵거나, 데이터 양의 부족—등 여러가지 이유에서 연구가 되어있지 않다. 이런 관점에서 이 연구 제안의 가장 큰 강점은 무엇보다도 실제로 채점을 하고 학생들과 교류하고 있는 조교들이 연구를 수행한다는 점이다. 이것은 이전의 다른 통계 연구와는 차별화 되는 강점이다. 또한 다양한 학생들이 조교를 거쳐간다는 점에서 미래의 풍부한 데이터가 이미 확보되어 있다.

**Novelty**

물론 네트워크상에서 정보 이동에 관한 연구가 이루어져 있고, 학술지 표절을 찾아내는 프로그램도 이미 사용되고 있다. 하지만 네트워크 분석과 통계물리학을 이용하여 학생들간 숙제 베끼기를 적발하기 용이하도록 고안된 프로그램은 아직 없다. 더욱이 숙제간의 네트워크는 어떤 프로그램도 말해주지 않는다. 이러한 프로그램을 개발한다면 교육적 측면에서 커다란 도움이 될 것이라 생각한다. 따라서 본 연구가 숙제들 간의 관계를 물리적, 수학적 방법론으로 연구한 최초의 사례가 될 것이다.

Scholarly profundity

**네트워크 공간의 측도 개발**

우리는 학생들의 숙제들 사이의 관계를 이해하는 방법으로 수학적으로 의미있는 공간을 정의하고 그것의 구조에 관한 연구를 하려한다. 숙제의 난이도, 학생들 사이의 거리, 채점하는 조교와

교수님의 성향 등 추상적인 공간으로부터 명확한 자료를 얻기 위해서는 수학적으로 엄밀하게 정의된 측도(measure)를 정의해야 한다. 멱급수 분포를 이용한 네트워크 구조분석 등, 이미 네트워크에 관한 측도 개발은 어느정도 이루어져 있다. 하지만 우리가 이해하고자 하는 속제 공간의 네트워크를 분석하기에는 주어진 시스템의 다양한 변수 혹은 정보들을 포함하지 못하고 있다. 따라서 본 연구의 결과로서 제안할 측도는 이전의 네트워크 이론에서 제안한 측도보다 더 많은 정보를 포함할 수 있어야 한다. 또한 연구 성과로서 새로이 개발될 측도는 다양한 정보를 포함한 네트워크에서 동역학을 기술하는데도 쓰일 수 있다는 점에서도 앞으로도 네트워크 이론에서 응용될 가능성이 충분히 높다.

### **상전이 현상으로서의 속제 베끼기**

학생들의 속제 베끼기는 통계물리에서의 상전이와 비교할 수 있다. 특히 여과 전이(percolation transition)와 비교할 수 있을 것이다. 여과 전이는 격자공간에서 각 격자점, 혹은 연결선에 도체와 부도체가 존재할 확률이 매개 변수  $p$ 인 베르누이 분포를 따를 때, 매개 변수와 어떤 값  $p_c$ 의 대소관계에 의해 전체 시스템의 도체-부도체 상전이가 일어난다. 이 과정에서 초기에는 작은 클러스터가 형성되지만, 매개 변수가 증가함에 따라 클러스터가 생성, 성장, 병합되는 것을 관찰할 수 있다. 속제 베끼기의 경우에도 학생들을 점으로 생각하고 각 학생들 사이의 연결을 생각하고 특정 조건 이상에서 속제 베끼기 클러스터가 형성되고 그 이하에서는 그러지 않을 것이므로 생각할 수 있으므로, 이를 이용해 학생들의 속제 베끼기를 여과 전이로서 이해할 수 있을 것으로 생각한다.

또한, 이 뿐만이 아니라 초유체에서의 양자소용돌이의 형성 등, 다른 상전이 현상과도 비교할 수 있다. 초유체에서는, 엔트로피와 에너지가 서로 경쟁하며, 이를 통해 양자소용돌이가 생성되기도 하고, 소멸되기도 하며, 소용돌이에 할당되어 있는 전하가 커지거나 작아지거나 하고, 이는 계의 자유에너지를 통해 나타내어 진다. 속제 베끼기의 경우에도, 문제의 난이도, 문제 수 등의 요인들이 서로 경쟁하며 속제를 베끼지 말지가 결정되므로, 초유체에서의 양자소용돌이와 유사하게 생각할 수 있다. 따라서 여러가지 현상론적 논의와 유사성을 통해 속제 베끼기에 대한 라그랑지안이나 해밀토니안, 자유에너지 등을 얻을 수 있을 것이고, 이를 이용하여 좀 더 이론적인 접근 및 해석이 가능할 것으로 생각된다. 더 나아가 사전에 필요한 변수들을 알고 있을 경우, 연구를 통해 얻어진 운동방정식으로 결과를 예측할 수 있을 것이다. 혹은 결과를 이용하여 초기에 학생들 사이의 어떠한 조건이 조성되어있었는지도 역으로 추적할 수 있을 것이다.

### **의사 결정과의 유사성**

본 연구는 의사 결정 과정을 물리적으로 해석하는 데에도 응용이 가능할 것이다. 특정한 의제에 대하여 사람들이 쉽게 판단을 내릴 수 없는 경우, 외부의 출처(reference)에 의존하거나 뛰어난 사람의 의견을 좇을 텐데, 이것은 어려운 속제를 잘하는 다른 학생의

	<p>숙제를 참고하거나 출처를 찾아 참고해 푸는 것과 대응시킬 수 있을 것이다. 이러한 대응관계를 통하여, 우리의 연구를 의사 결정 과정에 응용할 수 있다. 뿐만 아니라, 선거에서 어느 후보에게 투표할지와 연관시켜 당선 후보 예측 등에도 이용될 수 있을 것으로 생각된다.</p>
<p>Research plan</p>	<p><b>5~6월</b></p> <p>학생들이 제출한 숙제의 정보를 전산화. 가장 중요한 일은 꾸준히 데이터를 쌓아 나가기 위해 숙제채점 과정과 결과를 잘 기록해두는 것이다. 이 과정에서 엄청난 양의 데이터가 쌓여가기 때문에 저장해 둘 곳과 분석장치를 미리 마련해 놓는 것이 중요할 것이라 예상한다. 그와 동시에 가능하다면, 수학적 그리고 물리적 기반으로 어떤 양을 측정해야할 지를 알려주는 측도를 개발하고자 한다.</p> <p><b>7~8월</b></p> <p>한 학기 동안 취합한 숙제를 가지고 1차 데이터 분석 및 적절한 통계역학적 함수 찾기. 봄학기 동안 쌓인 데이터를 기반으로 모델을 제안한 뒤 여름 방학 동안 데이터를 분석하고 부족했던 점들을 공부하고, 그 점들을 가을학기 데이터 수집에 반영한다. 누적된 결과를 통해 여름방학에 제안된 모델을 수정, 보완한 뒤 통계물리적, 수학적 아이디어를 통해 분석할 예정이다.</p> <p><b>9~10월</b></p> <p>새로운 과목에서 새로운 학생들이 제출한 숙제의 정보를 지속적으로 전산화한다. 1학기 후반에 진행했던 숙제의 전산화 작업을 새 학기에 출제된 다른 숙제에 대하여 다시 진행한다.</p> <p><b>11~12월</b></p> <p>가을 학기 숙제 정보를 이용한 1차 분석과 통계물리 함수찾기를 통해 적절하게 찾은 함수인지 검증. 수학적, 물리적 직관을 통해 제안된 모델이 상전이 현상을 볼 수 있는 모델이라 예상하고 있으므로 그것이 잘 반영 되었는지, 그 모델로 어떤 물리, 수학 혹은 사회현상을 기술하고 예측할 수 있는지 공부해 볼 수 있다.</p>