Lecture 10 Script

10.1 Convergence of Graph Filters in the Spectral Domain

Slide 1: Convergence of Graph Filters - Title Page

1. We consider convergent sequences of graphs along with associated sequences of graph filters. We show that if the graph sequences converge towards a graphon, the graph filter sequence converges towards a graphon filter in the frequency domain.

Slide 2: Graphon Filters and Sequences of Graph Filters

1. We are given a set of filter coefficients h_k

2. Which we use to construct a sequence of graph filters that are defined over a sequence of graphs. For each shift operators S_n in the graph sequence, we have a filter H-of-S_n in the graph filter sequence. Defined as the usual polynomial with coefficients h_k.

3. We also use the **same** coefficients to build a graphon filter. Which is a polynomial with the same coefficients but one in which the variable is the graphon shift operator T_W.

4. The question we address here is" Does the graph filter converge to the graphon filter. This is a good question and we are going to provide a response. But it is not the most pertinent question

5. Filters are used to process signals. Thus, filter convergence is important only inasmuch as it implies convergence of filter outputs. A more pertinent question is, consequently, whether graph filter outputs converge to graphon filter outputs.

6. To study this question we consider a convergent sequence of graph signals G_n-x_n that converge towards the graphon signal W-X.

7. The graph signal x_n is then fed as an input to the graph filter H-of-S_n to produce the output graph signal y_n. We do this for all the signal-graph pairs in the sequence.

8. Likewise, we feed the graphon signal X to the graphon filter T_H to produce the output graphon signal Y.

9. Our objective here is to describe convergence results for the graph signal sequence G_n-y-n. This sequence can be shown to converge to the graphon signal W-Y under some conditions. Thus taking the limit of a graph signal sequence and applying a filter is the same as applying a filter and taking the limit. We can understand graph filters in the limit by studying the corresponding graphon filter.

Slide 3: Graph filters, Graphon Filters, and Their Frequency Representations

1. Underlying this convergence analysis is the algebraic equivalency between graph filters, graphon filters, and their respective frequency representations. More formally, when we are given filter coefficients h_k, we have five different polynomials, which are all the same except that they are evaluated at different variables.

2. Two of these polynomials are representations in the node domain.

3. They are the graph filter sequence H-of-S_n which is a polynomial on variable S_n.

4. And the graphon filter T_h which is the same polynomial defined in variable T_W.

5. The other three polynomials are representations in the spectral domain.

6. The most important of the three is the frequency response of the graph and graphon filters. This is a polynomial on the scalar variable lambda. We recall the very important fact that the frequency responses of the graph filter and the graphon filter are the same. If you care to know, this is because a graph filter and graphon fitler are the same algebraic object.

7. Evaluating the frequency response of the graph at the graph eigenvalues gives the frequency representation of the graph filter.

8. And evaluating the frequency response of the graphon, which is the same response of the graph filter, at the graphon eigenvalues, gives the frequency representation of the graphon filter.

Slide 4: Convergence of Graph Filter Sequences in the Frequency Domain

1. Having introduced these preliminary definitions we can present a theorem that claims convergence of the graph filter sequences in the frequency domain.

2. To state the theorem recall the definition of the frequency representation of the graph filter.

3. As well as definition of the frequency representation of the graphon filter. We have just introduced these two. They are the same polynomial evaluated at different frequencies.

4. Consider then the filter coefficients h_k as fixed and given.

5. These coefficients generate a sequence of graph filters H-of-S_n supported on a sequence of graphs G_n. All of the filters use the same coefficients. But they run on different graphs.

6. We use the same filter coefficients to generate the graphon filter T_H supported on the graphon W.

7. If the graph sequence G_n converges to the graphon W

8. The frequency representations of the graph filters converge to the frequency representation of the graphon

Slide 5: Frequency Domain Convergence is Not Much

1. The theorem says that the GFT representation of the graph filter sequence converges to the WFT representation of the graphon filter. Having convergence of frequency representations is neither unexpected nor strong.

2. That the result is not unexpected is due to the fact that the frequency **representations** converge because the eigenvalues converge and the frequency **responses** of the graph and graphon filters are the same.

3. As we grow the index n, the eigenvalues of the graph approach the eigenvalues of the graphon. Given that the frequency response is continuous, it is a polynomial, the responses converge as well. This is, literally, the proof of the theorem.

4. That the result in not strong is due to the fact that the GFT and the WFT are representations that exist in our minds. We use them for analyses. But we never

compute them. Filters operate in the node domain. Meaningful convergence results must make claims on signal values. Not on their transforms.

10.2 Convergence of Graph Filters in the Node Domain

Slide 6: Convergence of Graph Filters - Title Page

1. As we did in the previous section, we consider convergent sequences of graphs along with associated sequences of graph filters. We show that if the graph sequences converge towards a graphon, the graph filter sequence converges towards a graphon filter. But our statements are now in the **node** domain. Not the spectral domain

2. This convergence result provides our first approach to the transferability of graph filters.

Slide 6: From Frequency Representations to Node Representations

1. But the result is not useless either because we can leverage it to prove convergence in the node domain. To do that, we have to go into the frequency domain and back.

2. Namely, we begin with an input graph signal x_n and compute its GFT to represent it in the GFT domain.

3. We then process the GFT with the frequency representation of the graph filter.

4. And come back to the node domain by applying the inverse GFT.

5. Likewise, we get an input graphon signal X and compute its WFT to represent it in the WFT domain.

6. We then process the WFT with the frequency representation of the graphon filter.

7. And come back to the node domain by applying the inverse WFT.

8. This is a promising strategy because we know that the frequency representations converge. The highlighted blocks are the same in the limit.

9. However, the GFT and the inverse GFT do not converge to the WFT and inverse WFT. This proof strategy fails in general.

10. Except, of course, when the signals are graphon bandlitted. In this case the GFT and iGFT **do** converge to the WFT and iWFT.

Slide 7: Graph Filter Convergence for Bandlimited Inputs

1. We have therefore just proved a theorem claiming the convergence of graph filters when given bandlimited inputs.

2. To state the theorem formally, consider a sequence of graph signals G_n-x_n. This input sequence generates a sequence of output graph signals G_n-y-n. In which each of the output signals y_n follows from application of the graph filter H-of-S_n.

3. We also consider a graphon signals W-X. This input signal generates the output graphon signal W-Y. In which Y follows from application of the graphon filter T_H.

4. The sequence of input graph signals G_n-x_n is convergent. It converges to the graphon signal W-X.

5. The filters H-of-S_n and T_H are generated by the same filter coefficients h_k.

6. It then follows that if the input signals are c-bandlimited.

7. The sequence of output graph signals converges to the output graphon signal.

Slide 8: Lipschitz Graphon Filters

1. This theorem has been ready to prove. And it is also somewhat weak. Although we consider signals that, technically, have an increasing number of components, the signals are required to admit a representation on a finite number of components. Their spectral representations have a finite number of nonzero entries. A stronger result is desirable. And, as it turns out, also possible.

2. To do that we introduce the notion of a Lipschitz graphon filter. The definition is what you would expect. A graphon filter is Lipschitz if its frequency response is Lipschitz in the interval [-1,1].

3. That is, the absolute value of the difference between the filter's frequency response evaluated at lambda_1 and the filter's response evaluated at lambda_2.

4. Is bounded by the absolute value of the difference between lambda_1 and lambda_2, scaled by a Lipschitz constant L.

5. This condition must hold for any pair of variables in the [-1,1] interval.

6. If restrict our attention to filters that have Lipschitz frequency responses, we can claim convergence of the filter sequence even though we lack convergence if the GFT and iGFT sequences.

Slide 9: Graph Filter Convergence for Lipschitz Graphon Filters

1. Formally.

2. We are given a sequence of input graph signals G_n-x_n that converges to the graphon signal W-X.

3. Along with filters H-of-S_n and T_H that are generated by the **same** filter coefficients h_k.

4. It then follows that if the filter coefficients are associated with a frequency response that is Lipschitz.

5. The sequence of output graph signals converges to the output graphon signal.

6. The proof of this result is not cheap. You can find the proof in lecture notes that are accessible through the course's website. We point out that the theorem does not impose restrictions on the input signals as we did when claiming convergence for c-bandlimited inputs. The restriction in on the type of filters that are admissible for the theorem to hold. They have to have a frequency response that is Lipschitz.

Slide 10: Remarks on the Proof of Convergence for Lipschitz Graphon Filters

1. The theorem is, not surprisingly, proven in the spectral domain. The first critical observation in the proof is that for GFT and WFT components that are associated with eigenvalues that are not close to zero, convergence is ready. This is actually our theorem for c-bandlimited inputs.

2. The challenge arises when we focus on the components that are associated with eigenvalues that are not close to zero. We have seen that eigenvalues accumulate at zero.

3. And that this causes complications with eigenvector convergence. It makes it impossible for us to claim uniform convergence.

4. The second critical observation in the proof is to realize that the Lipschitz continuity hypothesis renders this problem void.

5. Indeed, since the filter is not changing too quickly, all GFT or WFT components that are associated with eigenvalues close to zero, are multiplied by similar number. It doesn't matter that we can't tell them apart because the filter does not distinguish among them either.

6. Look at the proof to see how these arguments play out.

Slide 11: Remarks on the Convergence of Lipschitz Graphon Filters

1. Asides from commenting on the proof it is also important to comment on the implications of the theorem. A fundamental issue that we identify is that transferability is counter to discriminability.

2. If the filter converges, it must be Lipshcitz. If it is Lipchitz, its rate of variability is limited by the Lipschitz constant L. But as we approach lambda equals 0 we have closer eigenvalues. Which means we would need thinner filters to tell components apart. But these thinner filters take longer to converge towards the graphon limit. Thus we either have wide filters that converge quickly to the graphon limit. These filters transfer well but are not discriminative. Or we have thin filters that take a while to converge to the graphon. These filters are discriminative. But they do not transfer well. The challenge is, not coincidentally, the same challenge we encountered in our stability analyses of graph filters. Also not coincidentally, the challenge is inherent to graph filters. For as long as we use graph filters, transferability and discriminability are incompatible. The solution to this conundrum is, what else, to resort to the use of graph neural networks.

3. Before we go there, we must contend with the weakness of our characterization. The result we provided is just a limit.

4. We will work on transference bounds that hold for finite n.

10.3 Graphon Filters as Generative Models for Graph Filters

Slide 1: Graphon Filters as Generative Models for Graph Filters – Title Page

1. We saw that, for convergent graph sequences, graph filters converge asymptotically to graphon filters. That suggests the use of finite graph filters to approximate graphon filters. We now discuss the conditions under which graph filters can approximate graphon filters, and how good that approximation is for different values of n.

Slide 2: Graphon Filters are Generative Models for Graph Filters

1. In the last few lectures, we saw that, for convergent graph sequences, the eigenvalues of the graph converge to the eigenvalues of the graphon. Moreover, as the frequency response of both graph and graphon filters are polynomials instantiated on the eigenvalues of the graph or graphon, this implies that the frequency response of graph filters converge asymptotically to the frequency response of a graphon filter.
2. As the number of nodes in the graph, n, grows, the graph filters become more and more similar to the graphon filter.
3. That suggests that we can use graph filters as approximations for the graphon filter. Not asymptotically, but for a graph with a finite number of nodes n.
4. That's precisely what we discuss in this lecture. We present conditions under which graph filters can approximate graphon filters. And we also quantify how good that approximation is for different values of n.

Slide 3: Small Eigenvalues are Hard to Discriminate

1. A trouble we face when approximating graphon filters with graph filters is the fact that the eigenvalues of a graphon accumulate around zero.
2. That accumulation makes it hard to match graph eigenvalues to the corresponding graphon eigenvalues when lambda is small

Slide 4: Small Eigenvalues are Hard to Discriminate

1. Which in turn makes it hard to discriminate between consecutive eigenvalues in that range
2. Thus, if we allow the filter to change rapidly around zero, the filter may modify the graph and graphon eigenvalues differently.

3. To obtain good approximations of a graphon filter with a graph filter, we must then restrict our attention to filters that do not change much around lambda = 0

Slide 5: Low-Pass Lipschitz Filters

1. We start with low-pass Lipschitz Filters. As graphon eigenvalues tend to zero as the index grows,

2. Low-pass filters must be zero for all eigenvalues below a certain threshold c, with the constant c determining the filter's band.

3. The filter removes high-frequency components, that is, the eigenvalues of the graphon closer to zero. But low-frequency components are not affected. That is important because, on the one hand, low-frequency components are easier to match with graph eigenvalues. On the other hand, as zero is the only point of accumulation for graphon eigenvalues, the number of eigenvalues in the passing band will be finite, allowing us to derive bounds on how close the approximation of a graphon filter by a finite graph filter will be.

Slide 6: Assumptions

1. To derive those approximation bounds, we need to introduce a set of Lipschitz assumptions on the graphon, the filter and the graphon signal.

2. First, we require the graphon to be Lipschitz with Lipschitz constant L_1. That is, for all pairs of arguments (u 1, v 1) and (u 2, v 2), it holds that the absolute value of the difference of the graphon evaluated at (u 1, v 1) and of the graphon evaluated at (u 2, v 2) is bounded by the Lipschitz constant L1 times the sum of the absolute value of the difference between u2 and u1, and the absolute value of the difference between v2 and v1.

3. We also require the filter's frequency response to be Lipschitz with constant L_2. That is, for all eigenvalues lambda 1 and lambda 2, we have that the absolute value of the difference of the filter's response evaluated at lambda 2 and lambda 1 is bounded by the Lipschitz constant times the distance between the eigenvalues. We also require the filter's frequency response to be normalized, in the sense that the absolute value of the frequency response at any frequency lambda is at most 1.

4. Our third assumption is that the graphon signal X is Lipschitz with constant L_3. As usual, that means that, for all u1 and u2, the absolute value of the difference of the graphon signal at u1 and u2 is bounded by the Lipschitz constant L3 times the distance between the arguments u1 and u2

Slide 7: Definitions

1. We will also need to introduce some definitions so that we can compare graph filters to graphon filters.
2. These definitions require that we fix a bandwidth c to separate eigenvalues that are close to 0 from those that are not close to 0.
3. Associated to c, we define the c-band cardinality B_n-c. This is a count of the number of eigenvalues whose absolute value is larger than c. This is a number we know is finite.
4. Also associated with c we define the c-eigenvalue margin of the graph G_n. This margin is the smallest gap between a graph eigenvalue lambda_ni and a graphon eigenvalue lambda_j with **different index.** The graph eigenvalue has to be in the c-band. But the graphon eigenvalue can be anywhere. Typically, this gap is the difference between the graph eigenvalue that is immediately above c and the graphon eigenvalue that is immediately below c.

Slide 8: Graph-Graphon Filter Approximation Theorem for Low-Pass Lipschitz Filters

1. We are now ready to state our first result. It refers to the approximation of a graphon filter by a graph filter with Low-Pass Lipschitz filters
2. Consider then a graphon Y and a graph filter y_n instantiated from Y. With definitions D1-D2, assumptions A1-A3 and the additional requirement that
3. The filter is low-pass, that is, h of lambda is equal to zero for all eigenvalues with absolute value below the threshold c
4. The L2 norm of the difference between the graphon filter Y and the graph filter Yn induced by yn
5. Is bounded by
6. The square root of the graphon's Lipschitz constant, L1
7. Times the sum of the filter's Lipschitz constant, L2,
8. And the ratio between pi times the number of eigenvalues in the passing band, n c, and the c-eigenvalue margin of the graph
9. And then times n to minus ½, times the L2 norm of the graphon signal X
10. To which we add another term that is comprised of the graphon signal's Lipschitz constant, L3, divided by the square root of 3 and multiplied by n to minus ½

11. We do not discuss the proof here, but the proof is available on the course website.

Slide 9: High-Pass Filters

1. In order to see how we can deal with the graphon eigenvalues accumulated around zero, we now turn our attention to high-pass filters. High-pass graphon filters are those that have null frequency response for all eigenvalues that have absolute value above a certain threshold c
2. That is, high-pass filters remove components associated to low graphon frequencies
3. Here, we also assume that the filters have low variability around zero
4. That low variability of the filter in high frequencies makes it easier to match graph eigenvalues to graphon eigenvalues around lambda = 0, thus leading to approximating graph filters with tighter approximation bounds

Slide 10: Graph-Graphon Filter Approximation for High-Pass Filters

1. We are now ready to state our next result, concerning the approximation of graphon filters by graph filters for high-pass filters
2. Consider then a graphon Y and a graph filter y_n instantiated from Y. With definitions D1-D2, assumptions A1-A3 and the additional requirement that
3. The filter has low variability around zero and it is high-pass, that is, h of lambda is equal to zero for all eigenvalues with absolute value above the threshold c
4. The L2 norm of the difference between the graphon filter Y and the graph filter Yn induced by yn
5. Is bounded by
6. The filter's Lipschitz constant, L2, times the passing band threshold c, times the L2 norm of the graphon signal X
7. We do not discuss the proof here, but the proof is available on the course website.

Slide 11: Lipschitz Filters with Variable Band

1. With those two results in hand, we are ready to stablish approximation bounds for more general filters. The class of filters that we consider here corresponds to those filters that exhibit low-variability for those eigenvalues located below a certain threshold c. That is precisely the region where the eigenvalues of the graphon accumulate, and where we

may have trouble matching graph eigenvalues to the corresponding graphon eigenvalues

2. For those eigenvalues located above the threshold c, we require the filter to be Lipschitz, but the filter doesn't need to sacrifice variability in this region. The number of eigenvalues satisfying lambda greater than c is finite. And the eigenvalues do now accumulate in that region, making it easier to match graph and graphon eigenvalues
3. Now, as you may have suspected, the class of filters that we are discussing in this slide is nothing more than a composition of a low-pass filter
4. And a high-pass one
5. That construction is key to obtain our final result for the approximation of graphon filters by graph filters, that we present next

Slide 12 Graphon Filter Approximation Theorem for Lipschitz Filters with Variable Band

1. We now state our main result, that stablishes a bound for the approximation of a graphon filter by a graph filter
2. Consider then a graphon Y and a graph filter y_n instantiated from Y. With definitions D1-D2, assumptions A1-A3 and the additional requirement that
3. The filter exhibits low-variability for all lambda below a certain threshold c
4. The L2 norm of the difference between the graphon filter Y and the graph filter Yn induced by yn
5. Is bounded by
6. The square root of the graphon's Lipschitz constant, L1
7. Times the sum of the filter's Lipschitz constant, L2,
8. And the ratio between pi times the number of eigenvalues in the passing band, n c, and the c-eigenvalue margin of the graph
9. And then times n to minus ½, times the L2 norm of the graphon signal X
10. To which we add another term that is comprised of the graphon signal's Lipschitz constant, L3, divided by the square root of 3 and multiplied by n to minus ½
11. And another term consisting of the filter's Lipschitz constant, L2, times the passing band threshold c, times the L2 norm of the graphon signal X

Slide 13: Proof

1. As we discussed a few minutes ago, a filter with variable band can be seen as the sum of a low-pass, L2-Lipschitz filter with h1 (lambda) equal to zero for all lambda less than c
2. And a high-pass filter exhibiting low-variability for high frequency components, that is, for all lambda less than c, and a null frequency response for all eigenvalues outside of that range
3. As our resulting filter is the sum of those two filters, we can use a triangle inequality to compute the L2 norm of the difference between the graphon filter Y, and the graph approximation Yn. That is equal to the norm of the difference between the output of the graphon filter for a graphon signal X and the output of the induced graph filter. Now, as the resulting filter h is the sum of two filters h1 and h2, we can use the triangle inequality to break down that term between
4. The norm of the different between the output of the graphon filter and the induced graph filter for the low-pass Lipschitz filter, h1
5. And the norm of the difference between the output of the graphon filter and the induced graph filter for the high-pass Lipschitz filter, h2
6. But we already know how to bound the first term on the right-hand side. That is the approximation bound we obtained for low-pass Lipschitz filters
7. Similarly, we know that the second-term on the right-hand side can be bounded by the approximation bound we obtained for high-pass Lipschitz filters
8. Summing up the two bounds, we then prove our approximation theorem for Lipschitz filters with variable band

Slide 14: Parse the Bound

1. According to the theorem,
2. The difference between the graphon filter Y and the graph filter approximation Yn is upper-bounded,
3. With the approximation bound depending on the filter transferability constant, given by the square root of the graphon's Lipschitz constant L1 times the sum of the filter's Lipschitz constant A2 plus the ratio between the filter's parameters n_c and delta_nc.
4. And on the difference between the graphon signal X and the graphon signal X n induced by the graph signal x n

5. The bound also depends on the graphon via the graphon's Lipschitz constant L1. We could make the bound tighter by decreasing the value of L1
6. But L1 also affects the variability of the graphon
7. Note that the bound decreases asymptotically with n, as expected since we know graph filters converge asymptotically to graphon filters. Moreover, as n grows, the transferability constant dominates the bound. That implies that the quality of the approximation --- or how close we can make the graphon and the approximating graph filter --- depends strongly on the transferability constant.

Slide 15: Filter Responses Determines the Approximation Bound

1. According to the theorem, the approximation bound is dominated by the transferability constant as n grows
2. The transferability constant, in turn, depends on the filter parameters: the filter's Lipschitz constant L2, the number of eigenvalues in the passing band, n_c, and the c-eigenvalue margin of graph G_n
3. Both the filter's Lipschitz constant L2 and the filter's passing band --- which corresponds to the interval from c to 1 --- determine the variability of the filter's spectral response. The Lipschitz constant affects the sharpness of the filter: by increasing it, we can make it the filter sharper, and thus more discriminative. But that in turn makes the transferability bound less tight. The filter's discriminability depends also on the filter's passing band.
4. But the number of eigenvalues in the passing band has to be limited: we must have n_c less than the square root of n
5. This restriction on the number of eigenvalues in the passing band is necessary to assure that the eigenvalues of the graph converge asymptotically to those of the graphon. When that condition is satisfied, the c-eigenvalue margin of the graph, delta_c, is nonnull and converges to the minimum eigengap of the graphon for any eigenvalue in the passing band. Which ensures convergence of the eigenvalues, thus making the approximation tighter.

Slide 16: Discriminability-Approximation Trade-Off

1. From our discussion on bound of the approximation of a graphon filter by a graph filter, we identify a fundamental issue: good approximation bounds are counter to the filter's discriminability

2. To obtain tight approximation bounds, we need filters that do not change by much around zero
3. But in that case the filter is unable to discriminate components associated to eigenvalues close to zero
4. For larger graphs, though, that is less of an issue. Good approximation of a graphon filter by a graph filter requires the number of eigenvalues in the passing band, n_c, to be less than the square root of n
5. Thus, as n grows larger, we can afford a larger number of eigenvalues, n_c, in the passing band
6. That ends up improving discriminability of the filter without penalizing the filter’s approximation bound

10.4 Transferability of Graph Filters: Theorem

Slide 1: Transferability of Graph Filters: Theorem - Title Page

1. We study transferability of graph filters to show that graph filters are transferable across graphs that are drawn from a common graphon.

Slide 2: Comparing Graph Filters through their Generating Graphon Filter

1. The first point to empathize is that in the previous sections we have shown results that hint at transferability but we actually not proven transferability. We have, instead, proen that graph filters can be close to graphon filters under some conditions.
2. Indeed, our results look at a graph G_n with n nodes sampled from a graphon W
3. And claim that the graph filter H-of-S_n running on the graph G_n is close the graphon filter T_H running on the graphon W. If the filter coefficients are the same in both filters.
4. A legitimate transferability claim requires that we compare two different graphs, possibly having different numbers of nodes. And that we show the corresponding filters to be close.

5. Namely, we have to consider graphs G_n and G_m. Both sampled from the graphon W.

6. And have to show that the graph filters H-of-S_**n** and H-of-S_**m** are close.

7. It is easy to see that this is true given what we know about the similarity between both graph filters to the graphon.

8. We know that the H-of-S_**n** is close to the graphon filter T_H.

9. And we also know that the graph filter H-of-S_**m** is close to the graphon T_H.

10. From these two facts it follows that the graph filter H-of-S_**n** is close to the graph filter H-of-S_**m.**

11. This is just the triangle inequality applied to the comparison of the filters via their relative comparisons to the graphon.

Slide 3: Running the Same Filter on Different Graphs

1. To state a formal theorem we begin with a formal problem formulation. We are considering the graph signal S_n-x_n along with the graph signal S_m-x_m. Both of these signals are sampled from the graphon signal W-X. The samplings are different as they contain different numbers of samples.

2. We then consider a given set of filter coefficients h_k and process the signals on their respective graphs.

3. That is, we run the filter with coefficients h_k on the graph S_n to process the signal x_n. This is done with a polynomial on S_n modulated with coefficients h_k. This produces the signal y_n.

4. And we also run the filter with the **same** coefficients h_k on the graph S_m to process the signal x_m. This is done with the same polynomial but instantiated on the graph S_m. The output signal is y_m.

5. We want to compare the output signals y_n and y_m. Since they have different numbers of components, we do not compare them directly. Rather, we compare the induced graphon signals Y_n and Y_m.

Slide 4: Assumptions

1. Our transferability results require the same set of assumptions we used when comparing graph filters to graphon filters. This is a triplet of Lipschitz assumptions that we repeat here for reference.

2. We require the graphon to be Lipschitz with Lipschitz constant L_1.

3. The filter's frequency response to be LIpschitz with constant L_2.

4. And the graphon signal X to be LIpschitz with constant L_3. Of these three assumptions it is (A2), the LIpschitz continuity of the frequency response of the filter, that is fundamental. The other two assumptions are necessary. But they are not conceptually that important.

Slide 5: Definitions

1. We also require the same pair of definitions we introduced when comparing graph filters to graphon filters.

2. These definitions require that we fix a bandwidth c to separate eigenvalues that are close to 0 from those that are not close to 0.

3. Associated to c, we define the c-band cardinality B_n-c. This is a count of the number of eigenvalues whose absolute value is larger than c. This is a number we know is finite.

4. Also associated with c we define the c-eigenvalue margin of the graph G_n. This margin is the smallest gap between a graph eigenvalue lambda_ni and a graphon eigenvalue lambda_j with **different index.** The graph eigenvalue has to be in the c-band. But the graphon eigenvalue can be anywhere. Typically, this gap is the difference between the graph eigenvalue that is immediately above c and the graphon eigenvalue that is immediately below c.

Slide 6: Transferability Theorem

1. With setup, assumptions, and definitions out of the way, we can state the graph filter transferability theorem.

2. Consider then graph signals S_n-x_n and S_m-x_m sampled from the graphon signal W-X.

3. Further consider the corresponding filter outputs y_n and y_m. The outputs are the result of processing different signals on different graphs. But they both use the same set of filter coefficients.

4. If assumptions (A1)-(A3) hold. And considering definitions (D1) and (D2).

5. The difference between the norms of the induced graphon signals Y_n and Y_m.

6. Is bounded by the expression shown. There is a lot going on this bound.

7. We have terms that depend on the number of nodes of the graph. They are sums of inverse square roots and vanish as n and m grow.

8. The summands are scaled by the Lipschitz constants of the graphon.

9. The filter's frequency response.

10. And the graphon signal.

11. We also have terms that depend on the bandwidth c along with the corresponding c-band cardinalities and the c-eigenvalue margins of the graphs G_n and G_m. The bound depends on the larger of the two c-band cardinalities and the minimum of the two c-eigenvalue margins. We pick the worst of the two in each case. The one that makes the bound largest.

12. And we also have the norm of the graphon signal X appearing in the bound.

13. The proof of the theorem is a simple application of the triangle inequality. You can find it in lecture notes that are available on the course's website. More important than providing this proof, is to explore the effects of the different terms that appear in the transferability bound. We do this in the next section.

10.5 Transferability of Graph Filters: Remarks

Slide 7: Transferability of Graph Filters: Remarks - Title Page

1. In the previous section we showed that graph filters are transferable across graphs that are drawn from a common graphon. There are several important remarks that follow from this theorem. Which we cover in this section.

Slide 7: Thing 1, Thing 2, and Thing 3

1. Let's pay homage to Dr. Seuss and point out that we have three things in this bound.

2. Thing 1 is a term that comes from the discretization of the graphon signal. It is not very important. In the sense that it is something we should had expected to appear and that it has the effect we should had expected it to have. Graphon signals that vary more rapidly are more difficult to approximate with their samples.

3. Thing 2 is a term coming from the filter's variability at eigenvalues with absolute value larger than c. This is the part of the bound that is associated with the components for which convergence is "easy." By which we mean the eigenvalues that are not excessively close to other eigenvalues. This is the part of the bound that is most interesting and whose derivation is most difficult. It gives a clear picture of what is the error incurred in these frequency components when we move from one graph to another. This is the inherent transferability error of the filter.

4. Thing 3 is a term that comes from the filter's variability at eigenvalues with absolute value smaller than c. This is the part of the bound that is associated with the components for which convergence is "difficult." By which we mean the eigenvalues that are very close to other eigenvalues because they are clustered at lambda=0. This is a part of the bound that is a sort of residual error. They are analysis leftovers coming from the frequency components that are difficult to tell apart. It is, nevertheless, a part of the bound that is fundamental. It characterizes a part of the spectral representation that is impossible to transfer from one graph to another.

5. Thing 2 and Thing 3 are, therefore, inherently different. Thing 2 characterizes the error in the part of the spectrum that can be transferred from one graph to the other. Thing 3 is

the error associated with the part of the spectrum that we can't transfer from one graph to another.

6. (Empty)

Slide 8: All Filters are Transferable in the Limit

1. Continuing with our parsing of the transferability theorem, the next point to emphasize is that all filters are transferable in the limit.

2. Indeed, as we let n and m grow most of the transferability error decreases and eventually vanishes. This decrease is proportional to the number of nodes in the smaller graph, which is the one that dominates asymptotically. There is a residual error term that does not vanish. This is associated with the part of the spectrum clustered around lambda=0. The part of the spectrum that we can't transfer from one graph to the other.

3. It is interesting, however, that for larger graphs we can also afford a higher c. This is because larger n and m compensate for large c-band cardinalities and smaller c-eigenvalue margins. If we decrease the bandwidth c, the c-band cardinalities grow the c-eigenvalue margins decrease. But they are canceled out by the larger number of nodes in the graph.

4. We can thus reduce the last term. The residual error associated with small eigenvalues. Because we can cancel out the resulting increase of c-band cardinalities and the resulting decrease of c-eigenvalue margins.

5. It is also interesting to point out that larger number of nodes also compensates for increasing Lipschitz constant in the filters' frequency responses. This allows for using sharper filters in larger graphs. We can transfer more discriminative filters when the graph grows larger.

6. (Empty)

Slide 9: Rates of Change of Graphons and Graphon Signals

1. The rate of change of the graphon and the graphon signals has an effect on the bound. This is captured by the Lipschitz constants of the graphon and the graphon signal.

2. The effect is rather straightforward. Graph signals and graphons with larger variability make filter transference more difficult.

3. This is because of sampling approximation errors. It is not a fundamental aspect of the story.

4. We remark that this constants can be sharpened if we introduce Lipschitz constants modulo permutation. But since this part of the error is not fundamental. The exercise is rather pointless.

5. (Empty)

Slide 10: Filter Discriminability

1. The most important feature of the transferability bound is the effect of filter discriminability.

2. This is captured by the Lipschitz constant of the filter's frequency response. Whose growth produces a concomitant linear rise on the value of the transferability bound. Thus, filters that are more discriminative, are more difficult to transfer.

3. It is interesting to point out that this is true in the part of the bound related to the components that can be transferred. Those that are associated with eigenvalues that have absolute value larger than c.

4. And it is also true of the part of the bound that corresponds to components that cannot be transferred. Those that are associated with eigenvalues that have absolute value smaller than c.

5. A reduction on the Lipschitz constant L_2 reduces both of these terms. The error in the part of the filter that can be transferred decreases. And the error in the part of the filter that cannot be transferred decreases as well.

6. (Empty)

Slide 11: Spectral Properties of the Graphon

1. A feature of the transferability bound that is a close second in terms of importance, is the effect of the spectral properties of the graphon.

2. A first important point to make is that the bound is parametric on the bandwidth c. Different choices of c result in different values for the bound. There is a value of c that makes this bound smallest. This is because the term L_2-c decreases with c. But the ratio of the band cardinality and the eigenvalue margin increases with c

3. The second important point is that increases in the c-band cardinality or decreases in the c-eigenvalue margin result in more challenging transferability. This is something that should be expected. When the c-band cardinality increases there are more components to transfer. When the eigenvalue margin decreases, the eigenvectors are more difficult to separate.

4. A final point to make is that although we write the c-band cardinalityand the c-eigenvalue margin as properties of the graphs, they are properties of the grpahon in the limit. This is because the graph eigenvalues converge to the graphon eigenvalues as we grow the size of the graphs.

5. The reason why this effect is second in importance is that it is not under our control. The graphon is what it is and its spectral properties are what they are. This is different from the effect of the filter's discriminability. Which is something that we can control.

Slide 12: Transferability vs Discriminability Non-Tradeoff

1. It is gratifying to see our discriminability analyses bringing us to a familiar point: The emergence of a Non-Tradeoff.

2. Indeed, if we fix n and m in the transferability bound we see that transferability and discriminability are incompatible.

3. For us to be able to discriminate frequency components around lambda=0 we need to have a large Lipschitz constant L_2. This is because eigenvalues cluster around lambda=0. We need sharp filters. But making L_2 large renders the transferability bound useless. We can't claim transferability if the Lipschitz constant of the filter's frequency response is large.

4. The solution of this dilemma, also brings us to a familiar place: The introduction of graph neural networks.

10.6 Graphon Neural Networks and GNN Transferability

Slide 1: Title Page

1. In this part of the lecture, we define graphon neural networks and discuss their interpretation as generative models for graph neural networks.

2. We then use graphon neural networks to show that GNNs inherit the transferability properties of graph filters.

Slide 2: Graphon Neural Networks

1. In this lecture, we have already seen that graph filters are transferable between weighted graphs associated with a given graphon. As such, in the same way that GNNs inherit the invariance and stability properties of graph filters, we can expect them to inherit transferability.

2. In order to study whether GNNs are transferable, we will once again turn our attention to graphons and graphon signals. In particular, we will use graphon signal processing to define graphon neural networks or WNNs.

3. The WNN can be defined as a layered architecture where each layer composes a graphon convolution with parameters h and a nonlinearity sigma.

4. The equation on the slide describes the ell th layer of a WNN with L layers and F_l output features. The h_kl^fg are its learnable weights. Recall that T_w^k is the graphon diffusion sequence of length k, defined by successive applications of the integral operator with kernel W, or graphon shift operator. The WNN input, X_0, is the graphon signal X. The WNN output, Y, is X uppercase L, the output signal at the last layer of the WNN.

5. Just like a GNN, the graphon neural network can be represented succinctly as a map Phi. This map maps the graphon signal X to the graphon signal Y, and it is parametrized by the graphon W and the set calligraphic H, which groups the learnable parameters h_kl^fg for all layers and all features of the WNN.

Slide 3: WNNs as Generative Models for GNNs

1. Consider a graph signal lowercase x on a graph G with shift operator S, and let Phi be a GNN applied to this graph signal and parametrized by the coefficient set H and the graph shift operator S. In this GNN, we know that the coefficient set H does not depend on the graph G. Likewise, consider a graphon signal uppercase X on the graphon W. The WNN map Phi applied to this graphon signal and parametrized by the coefficient set H and the graphon W does not depend on the graphon either.

2. Therefore, GNNs and WNNs can share their learnable coefficients H, as shown in the diagram on the slide. Meaning that we can use WNNs to instantiate GNNs.

3. In other words, the WNN acts as a generative model for GNNs. In the same way that graphons and graphon signals are generative models for graphs and graph signals.

4. In particular, in our transferability analysis we will consider GNNs Phi on weighted graphs G_n and graph signals x_n instantiated from the WNN Phi on the graphon W and graphon signal X. To instantiate a GNN from a WNN, it suffices to instantiate the graph shift operator S_n and the graph signal x_n from the graphon W and graphon signal X. The parameter set H is shared. Recall that, to instantiate the graph S_n and the graph signal x_n, we first construct a regular n-partition of the unit interval; then, we evaluate the graphon and the graphon signal at the endpoints u_i of each interval of this partition.

Slide 4: Sampling a GNN from the WNN

1. Consider a graph signal S_n, x_n sampled from the graphon signal W, X

2. And let calligraphic H be the set of coefficients of a graphon neural network with L layers and F features per layer, except for the input features F_0 and output features F uppercase L, which are equal to 1

3. Then, run a WNN Y with coefficients H on the graphon W to process the graphon signal X

4. And, from this WNN, instantiate a GNN y_n with coefficients H on the graph S_n to process the graph signal x_n

5. Now, we want to compare Y and y_n. But this is not possible, because the graphon signal X is a function and the graph signal x_n is a vector. To solve this problem, we consider the graphon signal uppercase Y_n, which is the graphon signal induced by the graph signal lowercase y_n.

Slide 5: Assumptions

6. Our transferability results require the same set of assumptions we used when comparing graph filters to graphon filters. This is a triplet of Lipschitz assumptions that we repeat here for reference.

7. We require the graphon to be Lipschitz with Lipschitz constant L_1.

8. The filter's frequency response to be LIpschitz with constant L_2.

9. And the graphon signal X to be Lipschitz with constant L_3. Of these three assumptions it is (A2), the LIpschitz continuity of the frequency response of the filter, that is fundamental. The other two assumptions are necessary. But they are not conceptually that important.

10. In the case of WNNs and GNNs, we also need a fourth assumption regarding the nonlinearities sigma. The first requirement is that sigma be normalized Lipschitz, which is what we call Lipschitz continuous functions with constant 1. In other words, the absolute value of sigma of x minus sigma of y should never exceed the absolute values of the difference between x and y. The second requirement is that sigma be null at zero. Neither of these requirements are very restrictive, as they are satisfied by most conventional activation functions, such as the ReLU, the sigmoid and the hyperbolic tangent.

Slide 6: Definitions

1. We also require the same pair of definitions we introduced when comparing graph filters to graphon filters.

2. These definitions require that we fix a bandwidth c to separate eigenvalues that are close to 0 from those that are not close to 0.

3. Associated to c, we define the c-band cardinality B_n-c. This is a count of the number of eigenvalues whose absolute value is larger than c. This is a number we know is finite.

4. Also associated with c we define the c-eigenvalue margin of the graph G_n. This margin is the smallest gap between a graph eigenvalue lambda_ni and a graphon eigenvalue lambda_j with **different index.** The graph eigenvalue has to be in the c-band. But the graphon eigenvalue can be anywhere. Typically, this gap is the difference between the

graph eigenvalue that is immediately above c and the graphon eigenvalue that is immediately below c.

Slide 7: GNN-WNN Approximation

1. With the setup, assumptions, and definitions in place, we can state the GNN-WNN approximation theorem.
2. Consider the graph signal S_n-x_n sampled from the graphon signal W-X.
3. Further consider the corresponding GNN and WNN outputs y_n and Y. The outputs are the result of processing these signals on the graph and on the graphon. But they both use the same set of filter coefficients.
4. If assumptions (A1)-(A4) hold.
5. And considering definitions (D1) and (D2).
6. The difference between the norms of the induced graphon signal Y_n and the graphon signal Y.
7. Is bounded by the expression shown on the slide. There is a lot going on in this bound.
8. We have terms that depend on the number of nodes of the graph. They are sums of inverse square roots and vanish as n and m grow.
9. The summands are scaled by the Lipschitz constants of the graphon
10. The filter's frequency response.
11. And the graphon signal.
12. We also have terms that depend on the bandwidth c along with the corresponding c-band cardinality and the c-eigenvalue margin of the graph G_n.
13. Finally, the bound also depends on the WNN depth L and the width F.
14. And we also have the norm of the graphon signal X appearing in the bound.

15. The proof of the theorem is an extrapolation of the graph-graphon filter approximation theorem for multiple layers and features. You can find it in lecture notes that are available on the course's website.

Slide 8: From WNNs to GNN Transferability

1. The GNN-WNN approximation theorem tells us that the error incurred when using a GNN to approximate a WNN can be upper bounded.
2. And the approximation bound is practically the same as the one we obtain when approximating graphon filters with graph filters
3. With an additional dependence on the GNN depth L and width F. Meaning that deeper and wider WNNs are harder to approximate.
4. The GNN-WNN approximation theorem is important because it is the stepping stone to proving transferability of GNNs. This is because the distance between GNNs and a WNN can be combined to bound the distance between GNNs supported on weighted graphs with different number of nodes instantiated from a graphon
5. Consider the WNNs Y_n, induced by the GNN on the n-node graph G_n, and Y_m, induced by the GNN on the m-node graph G_m. The distance between Y_n and Y_m
6. Can be bounded by adding and subtracting the graphon signal Y
7. And splitting the sum between the error incurred when approximating the WNN with the GNN on n nodes, and the error incurred when approximating the WNN with the GNN on m nodes
8. The sum can be split in two by the triangle inequality.
9. Because the two error terms on the left hand side are bounded, we have proved GNN transferability from G_n to G_m

Slide 9: Running the Same GNN on Different Graphs

11. Let us formally set up the transferability problem for GNNs. Consider the graph signals S_n, x_n and S_m, x_m sampled from the graphon signal W, X

12. And let calligraphic H be the set of coefficients of a graph neural network with L layers and F features per layer, except for the input features F_0 and output features F uppercase L, which are equal to 1

13. Then, run a GNN y_n with coefficients H on the graphon G_n to process the graph signal x_n

14. And a GNN y_m with coefficients H on the graph G_m to process the graph signal x_m

15. Now, we want to compare y_n and y_M. But this is not possible, because these are graph signals, and therefore vectors, with different dimensions m and n. To solve this problem, we consider the induced graphon signals uppercase Y_n and uppercase Y_m

Slide 10: Transferability Theorem

16. With the same setup, assumptions, and definitions used in the GNN-WNN approximation theorem, we can now state the GNN transferability theorem
17. Consider the graph signals S_n-x_n and S_m-x_m sampled from the graphon signal W-X.
18. Further consider the corresponding GNN outputs y_n and y_m. The outputs are the result of processing these signals on the graphs G_n and G_m. But they both use the same set of filter coefficients.
19. If assumptions (A1)-(A4) hold.
20. And considering definitions (D1) and (D2).
21. The difference between the norms of the induced graphon signals Y_n and Y_m
22. Is bounded by the expression shown on the slide. There is a lot going on in this bound.
23. Like before, we have terms that depend on the number of nodes of the graph. They are sums of inverse square roots and vanish as n and m grow.
24. The summands are scaled by the Lipschitz constants of the graphon
25. The filter's frequency response.
26. And the graphon signal.
27. We also have terms that depend on the bandwidth c along with the corresponding c-band cardinalities and the c-eigenvalue margins of the graphs G_n and G_m. The bound depends on the larger of the two c-band cardinalities and the minimum of the two c-eigenvalue margins. We pick the worst of the two in each case. The one that makes the bound largest.

28. Finally, the bound also depends on the WNN depth L and the width F.

29. And we also have the norm of the graphon signal X appearing in the bound.

30. The same comments made for the graph filter transferability theorem apply here, as the bound is almost exactly the same.

31. The only difference is the dependence on the depth L and width F of the GNN.

Slide 11: Transferability-Discriminability Tradeoff for GNNs

1. Because the GNN transferability bound looks almost exactly like the graph filter transferability bound, one might be inclined to think that graph filters and GNNs present the same discriminability-transferability non-tradeoff.

2. But in the case of GNNs, this non-tradeoff actually becomes a tradeoff. The nonlinearities help.

3. At each layer of the GNN, after the graph convolution, the nonlinearities sigma effectively scatter the eigenvalues of the graph around the spectra. In doing so, they allow some of the eigenvalues in the c-band

4. To be scattered to the upper and lower ends of the spectrum corresponding to eigenvalues with magnitude larger than c. Where these eigenvalues can later be discriminated by the graph convolutions of the upcoming layer.

5. The nonlinearities thus alleviate the discriminability-transferability trade-off. They give us room to decrease the c-band and increase the filter variability L_2, allowing to increase discriminability while retaining transferability.

6. This effect of nonlinearities is not reflected in the GNN transferability bound we have derived --- which is almost the same as the graph filter transferability bound. But it makes it so that, for the same level of discriminability, GNNs are more transferable than graph filters.